


```
;; PRIOR APPLICATION NUMBER: PCT/US01/00670
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: US 60/234,687
;; PRIOR FILING DATE: 2000-09-21
;; PRIOR APPLICATION NUMBER: US 09/608,408
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/774,203
;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 34947
;; LENGTH: 188
;; TYPE: prt
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AP000163.1
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.4
;; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.4
;; OTHER INFORMATION: EXPRESSED IN HEAT, SIGNAL = 1.7
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.92
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.1
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.5
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.5
;; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.95
;; OTHER INFORMATION: EST_HUMAN HIT: BE831331.1, EVALUATE 4.00e-65
;; OTHER INFORMATION: SWISSPROT HIT: P11308, EVALUATE 8.00e-99
US-09-864-761-34947
```

```
Query Match 38.8%; Score 948; DB 10; Length 188;
Best Local Similarity 95.5%; Pred. No. 1.4e-71;
Matches 169; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
```

```
OY 275 LANPGSGOIQIOMQFLELLSDSSNSCTIEGNGEFGKMTDPEVARRMGERSKPMNY 334
| : |||||||
DB 12 LSLSGGQIQIOMQFLELLSDSSNSCTIEGNGEFGKMTDPEVARRMGERSKPMNY 71
OY 335 DKLSRALRYRYDKNIMTKYGRKRYAKPDPHGIAQALQPHPESSMYKYPSDLPYMSYH 394
| : |||||||
DB 72 DKLSRALRYRYDKNIMTKYGRKRYAKPDPHGIAQALQPHPESSLYKYPSDLPYMSYH 131
OY 395 AHPQKNFVAHPHPALPYVSSSFFAAPNPYNSPTGIGIYPTNLPLPAHPSHIGTY 451
| : |||||||
DB 132 AHPQKNFVAHPHPALPYVSSSFFAAPNPYNSPTGIGIYPTNLPLSHMSHIGTY 188
```

RESULT 2

```
US-09-864-761-36929
;; Sequence 36929, Application US/09864761
;; Patent No. US20020048763A1
;; GENERAL INFORMATION:
;; APPLICANT: Penn, Sharon G.
;; APPLICANT: Rank, David R.
;; APPLICANT: Harel, David K.
;; APPLICANT: Chen, Wensheng
;; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;; FILE REFERENCE: Aegolica-X-1
;; CURRENT APPLICATION NUMBER: US/09/864,761
;; CURRENT FILING DATE: 2001-05-23
;; PRIOR APPLICATION NUMBER: US 60/180,312
;; PRIOR FILING DATE: 2000-02-04
;; PRIOR APPLICATION NUMBER: US 60/207,456
;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: US 09/632,366
;; PRIOR FILING DATE: 2000-08-03
;; PRIOR APPLICATION NUMBER: GB 24263.6
;; PRIOR FILING DATE: 2000-10-04
;; PRIOR APPLICATION NUMBER: US 60/236,359
;; PRIOR FILING DATE: 2000-09-27
;; PRIOR APPLICATION NUMBER: PCT/US01/00666
;; PRIOR FILING DATE: 2001-01-30
```

```
;; PRIOR APPLICATION NUMBER: PCT/US01/00667
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00664
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00669
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00665
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00668
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00663
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00662
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00661
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00670
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: US 60/234,687
;; PRIOR FILING DATE: 2000-09-21
;; PRIOR APPLICATION NUMBER: US 09/608,408
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/774,203
;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 36929
;; LENGTH: 144
;; TYPE: prt
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AP000021.2
;; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.2
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.97
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.5
;; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.4
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1
;; OTHER INFORMATION: SWISSPROT HIT: P11308, EVALUATE 1.00e-87
;; OTHER INFORMATION: EST_HUMAN HIT: BE831331.1, EVALUATE 1.00e-59
US-09-864-761-36929
```

```
Query Match 32.5%; Score 794; DB 10; Length 144;
Best Local Similarity 97.9%; Pred. No. 5.6e-59;
Matches 141; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
```

```
OY 292 LLSDSNSCTIEGNGEFGKMTDPEVARRMGERSKPMNYDKLSRALRYRYDKNIMT 351
| : |||||||
DB 1 LLSDSNSCTIEGNGEFGKMTDPEVARRMGERSKPMNYDKLSRALRYRYDKNIMT 60
OY 352 KYHGKRYAYKDPHGIAQALQPHPESSMYKYPSDLPYMSYHAPQKNFVAHPHPALP 411
| : |||||||
DB 61 KYHGKRYAYKDPHGIAQALQPHPESSLYKYPSDLPYMSYHAPQKNFVAHPHPALP 120
OY 412 VYSSSFFAAPNPYNSPTGIGIYPN 435
| : |||||||
DB 121 VYSSSFFAAPNPYNSPTGIGIYPN 144
```

RESULT 3

```
US-09-850-799-2
;; Sequence 2, Application US/09850799
;; Patent No. US20020090647A1
;; GENERAL INFORMATION:
;; APPLICANT: Denieris, Evan S.
;; APPLICANT: Tyodoro, Dmitry V.
;; APPLICANT: Hendricks, Timothy J.
;; TITLE OF INVENTION: Reagents and Methods for the Screening of Compounds
;; FILE REFERENCE: Case-03828
;; CURRENT APPLICATION NUMBER: US/09/850,799
```

[illegible]

RESULT 6
 US-10-108-605-129
 ; Sequence 129, Application US/10108605
 ; Patent NO. US20020160934A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Broadus, Julie
 ; APPLICANT: Stam, Lynn
 ; APPLICANT: Bachmann, Jane
 ; APPLICANT: Kachdar, Kim
 ; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES FROM DROSOPHILA MELANOGASTER THAT ENCODE
 ; FILE REFERENCE: 31133B
 ; TITLE OF INVENTION: PROTEINS ESSENTIAL FOR LARVAL VIABILITY AND USES THEREOF
 ; CURRENT APPLICATION NUMBER: US/10/108,605
 ; CURRENT FILING DATE: 2002-03-27
 ; PRIOR APPLICATION NUMBER: US 09/761,142

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 9, 2002, 16:48:22 ; Search time 20.8751 Seconds
(without alignments)
635,672 Million cell updates/sec

Title: US-09-902-772-2

Perfect score: 2444 1 MASTIKALSVSEDSLFPE.....IYPNRLPAHMPHSILGTYR 451

Sequence: BLOSUM62

Scoring table: Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Issued_Patents_AA:*

- 1: /cgn2_6/ptodata1/1aa/5a_COMB.pep:*
- 2: /cgn2_6/ptodata1/1aa/5b_COMB.pep:*
- 3: /cgn2_6/ptodata1/1aa/6a_COMB.pep:*
- 4: /cgn2_6/ptodata1/1aa/6b_COMB.pep:*
- 5: /cgn2_6/ptodata1/1aa/PCITUS.COMB.pep:*
- 6: /cgn2_6/ptodata1/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2444	100.0	451	4	US-08-878-177-2
2	2262.5	92.6	478	4	US-08-878-177-4
3	1624.5	66.5	452	2	US-08-343-443B-4
4	507	20.7	340	4	US-09-360-779-2
5	507	20.7	340	4	US-09-360-779-2
6	424.5	17.4	468	4	US-09-435-335-2
7	421.5	17.2	470	4	US-09-092-636-9
8	340.5	13.9	543	2	US-08-469-412A-7
9	340.5	13.9	543	4	US-09-021-715-7
10	340.5	13.9	548	2	US-08-469-412A-2
11	340.5	13.9	548	4	US-09-021-715-2
12	324	13.3	452	4	US-09-055-113-4
13	303	12.4	78	2	US-08-469-412A-11
14	303	12.4	78	4	US-09-021-715-11
15	301.5	12.3	109	2	US-08-343-443B-6
16	294.5	12.0	335	4	US-09-055-113-1
17	286	11.7	555	2	US-08-780-835B-2
18	286	11.7	555	4	US-09-303-268-2
19	286	11.7	555	4	US-09-116-049-2
20	283.5	11.6	462	3	US-08-875-844B-5
21	283.5	11.6	462	4	US-09-116-049-4
22	272.5	11.1	79	2	US-08-469-412A-12
23	272.5	11.1	79	4	US-09-021-715-12
24	270	11.0	78	2	US-08-469-412A-13
25	270	11.0	78	4	US-09-021-715-13
26	268	11.0	78	2	US-08-469-412A-10
27	268	11.0	78	4	US-09-021-715-10

28	261.5	10.7	300	3	US-09-009-913-5	Sequence 5, Appl
29	236.5	9.7	283	3	US-09-009-913-339	Sequence 339, App
30	231.5	9.5	255	3	US-09-009-913-9	Sequence 9, Appl
31	231.5	9.5	265	3	US-09-009-913-7	Sequence 7, Appl
32	211.5	8.7	521	1	US-08-368-281-2	Sequence 2, Appl
33	211.5	8.7	581	1	US-08-368-281-4	Sequence 4, Appl
34	208	8.5	371	1	US-08-746-789A-2	Sequence 2, Appl
35	202	8.3	114	4	US-09-055-113-3	Sequence 3, Appl
36	201.5	8.2	81	2	US-08-469-412A-14	Sequence 14, Appl
37	201.5	8.2	81	2	US-09-021-715-14	Sequence 14, Appl
38	194	7.9	80	2	US-08-469-412A-15	Sequence 15, Appl
39	194	7.9	80	4	US-09-021-715-15	Sequence 15, Appl
40	169	6.9	164	3	US-09-009-913-11	Sequence 11, Appl
41	116	4.7	1664	1	US-09-599-652-2	Sequence 2, Appl
42	116	4.7	1664	2	US-08-642-846-2	Sequence 2, Appl
43	116	4.7	1664	4	US-09-264-604-2	Sequence 2, Appl
44	110.5	4.5	81	2	US-08-469-412A-16	Sequence 16, Appl
45	110.5	4.5	81	4	US-09-021-715-16	Sequence 16, Appl

ALIGNMENTS

```
RESULT 1
US-08-878-177-2
; Sequence 2, Application US/08878177
; Patent No. 6294354
; GENERAL INFORMATION:
; APPLICANT: Chugai Selyaku Kabushiki Kaisha, 5001, Iwamoto et
; TITLE OF INVENTION: Cell Calcification Suppressing Proteins and Genes of
; FILE REFERENCE: Chugai Selyaku Kabushiki Kaisha 5001
; CURRENT APPLICATION NUMBER: US/08/878,177
; CURRENT FILING DATE: 1997-06-18
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 2
; LENGTH: 451
; TYPE: PRT
; ORGANISM: protein sequence from C-11 gene
US-08-878-177-2
```

Query Match 100.0%; Score 2444; DB 4; Length 451;
Best Local Similarity 100.0%; Pred. No. 4.9e-225;
Matches 451; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MASTIKALSVSEDSLFECAYGSPHLAKTEMASSSEYGGTSKMSPPRQDDMSQP	60
DB	1	MASTIKALSVSEDSLFECAYGSPHLAKTEMASSSEYGGTSKMSPPRQDDMSQP	60
QY	61	PARTYIMECPNPNVNGSRNSPDDCSYAKGKMYSSSDNVGMANTGYMEKHIPPMMTT	120
DB	61	PARTYIMECPNPNVNGSRNSPDDCSYAKGKMYSSSDNVGMANTGYMEKHIPPMMTT	120
QY	121	NERVIVPADFTLWSTDVHROWLEMAVKEYGLPDVDILLFONIDKELCKMTKXDFORLT	180
DB	121	NERVIVPADFTLWSTDVHROWLEMAVKEYGLPDVDILLFONIDKELCKMTKXDFORLT	180
QY	181	PSYADLLSLHLRLRGRATFIRPNTSYPEAQRITTRDPLYEDARRSAMTSHSHT	240
DB	181	PSYADLLSLHLRLRGRATFIRPNTSYPEAQRITTRDPLYEDARRSAMTSHSHT	240
QY	241	OSKATOPSSSTVPKTEOORPOLDPYQILGPTSSILANGSGQIQLOMFLLELLSDSSNS	300
DB	241	OSKATOPSSSTVPKTEOORPOLDPYQILGPTSSILANGSGQIQLOMFLLELLSDSSNS	300
QY	301	CTTWEGTNGEFKMTDPDEVARRMGERKSKPMNYDKLSRALRYYYDKNIMTKVGRAY	360
DB	301	CTTWEGTNGEFKMTDPDEVARRMGERKSKPMNYDKLSRALRYYYDKNIMTKVGRAY	360
QY	361	KFDHGTAQALQPPSPSSMTKTPSDLPYMSSTYAHQKMFVAPHPALPYTSSFFAA	420
DB	361	KFDHGTAQALQPPSPSSMTKTPSDLPYMSSTYAHQKMFVAPHPALPYTSSFFAA	420

QY 421 PNPYNSPTGGITPNTPLPAHMPHSLGTY 451
Db 421 PNPYNSPTGGITPNTPLPAHMPHSLGTY 451

RESULT 2

US-08-878-177-4
; Sequence 4, Application US/08878177
; Patent No. 6294354
; GENERAL INFORMATION:
; APPLICANT: Chugai Seiyaku Kabushiki Kaisha, 5001, Iwamoto et
; TITLE OF INVENTION: Cell Calcification Suppressing Proteins and Genes of
; FILE REFERENCE: Chugai Seiyaku Kabushiki Kaisha 5001
; CURRENT APPLICATION NUMBER: US/08/878,177
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.10
; SEQ ID NO 4
; LENGTH: 478
; TYPE: PRT
; ORGANISM: protein sequence from c-erb gene
US-08-878-177-4

Query Match 92.6%; Score 2262.5; DB 4; Length 478;
Best Local Similarity 86.3%; Pred. No. 1,2e-207;
Matches 430; Conservative 0; Mismatches 1; Indels 67; Gaps 3;

QY 1 MASTKEALSVSEDSLEFCAYGSPHLAKTEMTASSSEFGQTSKMSPPVPOQDWLSOP 60
Db 1 MASTKEALSVSEDSLEFCAYGSPHLAKTEMTASSSEFGQTSKMSPPVPOQDWLSOP 60
QY 61 PARVTIKMECPNQVNGSRNSPDSCSVAGKMWSSSDNVGMNGSYMEKHIPPNNMT 120
Db 61 PARVTIKMECPNQVNGSRNSPDSCSVAGKMWSSSDNVGMNGSYMEKHIPPNNMT 120
QY 121 NERRVIVPADPTLMSTDHVRQMLEAVKEYGLPDVILFQNTDGKELCKTKDDFORLT 180
Db 121 NERRVIVPADPTLMSTDHVRQMLEAVKEYGLPDVILFQNTDGKELCKTKDDFORLT 180
QY 181 PSYNADILSHLYRLER-----GATFIFPTSYPEA 213
Db 181 PSYNADILSHLYRLER-----GATFIFPTSYPEA 213
QY 214 TORITTRPDLPEQARRSAMTSHSPTQSKATQPSSTVPTKEDQRPQDLPYQIIGPTSS 273
Db 214 TORITTRPDLPEQARRSAMTSHSPTQSKATQPSSTVPTKEDQRPQDLPYQIIGPTSS 273
QY 241 TORITTRPDLPEQARRSAMTSHSPTQSKATQPSSTVPTKEDQRPQDLPYQIIGPTSS 300
Db 241 TORITTRPDLPEQARRSAMTSHSPTQSKATQPSSTVPTKEDQRPQDLPYQIIGPTSS 300
QY 274 RLANPSSGOIQLMQFLLELSDSSNSCITWEGTNGEFTKMTDPDEVARRMGERKSKPNMN 333
Db 301 RLANPSSGOIQLMQFLLELSDSSNSCITWEGTNGEFTKMTDPDEVARRMGERKSKPNMN 360
QY 334 YOKLSALRYYYDKNIMTVHGRKRAYKFDHGLAQLPHEPSSSMYVPSDLPYMSY 393
Db 361 YOKLSALRYYYDKNIMTVHGRKRAYKFDHGLAQLPHEPSSSMYVPSDLPYMSY 400
QY 394 -----HAHPQKMFVAPHPALPYTSSSFFAAPPYMNSPTGGIT 433
Db 401 HGRKRAYKFDHGLAQLPHEPQKMFVAPHPALPYTSSSFFAAPPYMNSPTGGIT 460
QY 434 PNTRLPAHMPHSLGTY 451
Db 461 PNTRLPAHMPHSLGTY 478

RESULT 3

US-08-343-443B-4
; Sequence 4, Application US/08343443B
; Patent No. 5968734
; GENERAL INFORMATION:
; APPLICANT: Aurias, Alain
; APPLICANT: Delattre, Olivier

APPLICANT: Desmaze, Chantal
APPLICANT: Melot, Thomas
APPLICANT: Peter, Martine
APPLICANT: Ploougaestel, Beatrice
APPLICANT: Thomas, Gilles
APPLICANT: Zuoman, Jessica
TITLE OF INVENTION: NUCLEIC ACID CORRESPONDING TO A GENE OF
TITLE OF INVENTION: CHROMOSOME 22 INVOLVED IN RECURRENT CHROMOSOMAL
TITLE OF INVENTION: TRANSLATIONS ASSOCIATED WITH THE DEVELOPMENT OF CANCEROUS
TITLE OF INVENTION: TUMORS, AND NUCLEIC ACIDS OF FUSION RESULTING FROM SAID
NUMBER OF SEQUENCES: 129
CORRESPONDENCE ADDRESS:
ADDRESSEE: Weiser & Associates
STREET: 230 South Fifteenth Street
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19102

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: AEDIT 1.0 DOS text editor
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/343,443B
FILING DATE: 18-NOV-1994
CLASSIFICATION: 514
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: PCT/FR93/00494
FILING DATE: 19-MAY-1993
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: FR 92/06123
FILING DATE: 20-MAY-1992
ATTORNEY/AGENT INFORMATION:
NAME: Weiser, Gerard J.
REGISTRATION NUMBER: 19,763
REFERENCE/DOCKET NUMBER: 989.6121P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-875-8383
TELEFAX: 215-875-8394
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 452 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-343-443B-4

Query Match 66.5%; Score 1624.5; DB 2; Length 452;
Best Local Similarity 67.0%; Pred. No. 9.5e-147;
Matches 306; Conservative 57; Mismatches 83; Indels 11; Gaps 7;

QY 1 MASTKEALSVSEDSLEFCAYG-SPHLAKTEMTASSSEFGQTSKMSPPVPOQDWLSO 59
Db 1 MASTKEALSVSDSLSDSAVAHLPKADMTASGSPDQGPQKINPLPQDQWLNQ 60
QY 60 PARVTIKMECPNQVNGSRNSPDSCSVAGKMWSSSDNVGMNGSYMEKH-IPPNM 118
Db 61 -PVRVAVKREY--DHNGRESVPDSCSVKSKLVGGESNPNNNTSYDEKNGPPNN 117
QY 119 TNERRVIVPADPTLMSTDHVRQMLEAVKEYGLPDVILFQNTDGKELCKTKDDFOR 178
Db 118 TNERRVIVPADPTLMSTDHVRQMLEAVKEYGLPDVILFQNTDGKELCKTKDDFOR 177
QY 179 LPPSYNADILSHLYRLERGAFTFPPNTSVPEARQRTTRPDLPEQARRSAMTSHSH 238
Db 178 ATTLINTEVLLSLTLRE--SSLAINTTSHDOSLSLVEDPSYDSVRRGAMGNNN 235
QY 239 PTQSKATQ-PSSSTVPTKEDQRPQDLPYQIIGPTSSRLANPSSGOIQLMQFLLELSDS 297
Db 236 SGLNKPPLGAGQTIKNTQRRPQDLPYQIIGPTSSRLANPSSGOIQLMQFLLELSDS 295

OY	298	NSNCITMGTGEEKMDPDVVARWBERKSKPMNNDKLSRLALRYXDKINIMTKHGKR	357
Dd	296	NASCIITMGITGEFKMTDPDVARWBERKSKPMNNDKLSRLALRYXDKINIMTKHGKR	355
OY	358	YAYKFDFHIGIQAOPHPRESSMKYPSPDI.PYMNSYAHBQKNMFVAHPALPYTSSP	417
Dd	356	YAYKFDFHIGIQAOPHPRESSMKYPSDSIYMPSTYAHBQKNMFVAHPALPYTSSSF	415
OY	418	FAAPNPYWNSPTGCTIPN---TRLPAAHMSLCTGY	451
Dd	416	FCAASQYTSTPGCTIPNPVPNPBPNTNHVSHSLCSTY	452

```

RESULT 4
US-09-360-779-2
Sequence 2, Application US/09360779
Patent No. 6268216
GENERAL INFORMATION:
APPLICANT: Denaris, Evan S.
APPLICANT: Evodoro, Dmitry V.
APPLICANT: Hendricks, Timothy J.
TITLE OF INVENTION: Reagents and Methods for the Screening of Compounds
TITLE OF INVENTION: Useful in the Treatment of Neurological Diseases
FILE REFERENCE: CASE-03828
CURRENT APPLICATION NUMBER: US/09/360,779
CURRENT FILING DATE: 1999-07-26
EARLIER APPLICATION NUMBER: 60/094,264
EARLIER FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 13
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 340
TYPE: PRF
ORGANISM: Rattus norvegicus
US-09-360-779-2

```

```

Query Match      20.7%; Score 507; DB 4; Length 340;
Best Local Similarity 49.8%; Pred. No.3,4e+40;
Matches 109; Conservative 18; Mismatches 40; Indels 52; Gaps

QY      269  GPTSSRLANPSCQIQIOWQFLLELLSDSSNSNCITWEGTNGEFGKMTDPEVARRMGERKS 328
      || : ||||| ||||| ||||| : || : ||||| ||||| |||||
Db      137  GPLSNV-QKSSGQQLWQFLLELLADANAGCIMEGHGSGFKLTDPDEVARRMGERKS 195

QY      329  KPNMNYDKLSRLRLRYDYDKNIMTKVHGKRAYKPDFHGIAQALQHPH----- 375
      KPNMNYDKLSRLRLRYDYDKNIMSKVHGKRAYRFPDQIAGQACQPPRPAHAHAHAHAHAHA 255
Db      196  KPNMNYDKLSRLRLRYDYDKNIMSKVHGKRAYRFPDQIAGQACQPPRPAHAHAHAHAHAHA 255

QY      376  ---PESSWTKYKPSDL-----PYMSSTYAAHPQKMFVAPHPALPYVTSSEFFAAPMPYNS 427
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      256  AAADGALYKYPAGIAPLPFPGLS-----KLTNMAASAGVAPAGFS-----YWPFG 300

QY      428  P-----TGGIYVPTRLP-----AAHMPSHLGTY 451
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db      301  PNATAAAAATATALYPTPGIQPPPGFPGVAAASHLGCH 339

RESULT 5
US-09-435-335-2
: Sequence 2, Application US/09435335
: Patent No. 6384204
: GENERAL INFORMATION:
: APPLICANT: Denetis, Evan S.
: APPLICANT: Eydoro, Dmitry V.
: APPLICANT: Hendricks, Timothy J.
: TITLE OF INVENTION: Reagents and Methods for the Screening of Compounds
: FILE REFERENCE: CASE-04027
: CURRENT APPLICATION NUMBER: US/09/435,335
: CURRENT FILING DATE: 1999-11-05
: EARLIER APPLICATION NUMBER: 09/360,779
: EARLIER FILING DATE: 1999-07-26
: NUMBER OF SEQ ID NOS: 23

```

```

; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 340
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-09-435-335-2

```

Query Match	20.7%;	Score 507;	DB 4;	Length 340;
Best Local Similarity	49.8%;	Pred. No. 3.4e-40;		
Matches 109;	Conservative 18;	Mismatches 40;	Indels 52;	Gaps 77;

Qy	269	GPTSSRLANPSSGGIQLMQFLLELLISBSSNCSNCTTGWGTCNFEFMTOPDEYARWGGRKS	328
		: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :	
Db	137	GPLSAVA-QKSSGQIQLMQFLLELLADRAANAGCTAMWGCHGEFLTTPDEYARWGGRKS	1959
Qy	329	KPNMNYDKLSBALRYYYDDKNTMTKRVHVGKRAYKPKDFHGIAQLALPHR-----	3759
Db	196	KPNMNYDKLSBALRYYYDDKNTMSYHVGKRAYKRRDFGQLAQCPRAAHMAAAAAAAAAA	2555
Qy	376	---PESSMYKTPSDL-----PYMSSTYHAHPQKMFVAAPHPALFVYTSSTFFAANPYWNS	4277
		: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :	
Db	256	AAADGALYKLPAGIAPLPFPGLS-----KLNTMAASAGVAPAGFS-----YWPQ	3000
Qy	428	P-----TGCIYPNTRLP-----AAHMPSHLGGYY	451
		: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :	
Db	301	PNATATAATATLPTPEGLQPPCPGFGVAASHLGGYY	339

```

RESULT 6
US-09-092-636-9
: Sequence 9, Application US/09092636A
: Patent No. 6162641
: GENERAL INFORMATION:
: APPLICANT: Goldman, Daniel
: APPLICANT: Sapru, Mohan K.
: TITLE OF INVENTION: Neuregulin Response Element and Uses Therefor
: FILE REFERENCE: DMI-003
: CURRENT APPLICATION NUMBER: US/09/092,636A
: CURRENT FILING DATE: 1998-06-05
: NUMBER OF SEQ ID NOS: 11
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 9
: LENGTH: 468
: TYPE: PRT
: ORGANISM: Mus musculus
US-09-092-636-9

```

```

Query Match          17.4%: Score 424.5: DB 4: Length 468;
Best Local Similarity 31.2%: Pred. No. 4.3e-32;
Matches 110; Conservative 34; Mismatches 93; Indels 115; Gaps 10.

QY 122 ERRVIVPADPILWSTDHVRQWLEMAVKEYGLDPVDILLFQNLIDGELCKMTKDFQRLTP 181
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db 94 QRRLGIPKPNPMLWSEQOYQWMLLMATNEFSLVNVLHOF-GANNQGLCMIGKERELELAP 152

QY 182 SYNADILLSHL-----HYREKGAIFI-----F 204
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db 153 DFVGDIILWHLLEQMKEKDEKTEDOYEENSHLNAVPHMINSTLGLSMEQAPYGMQAPNY 212

QY 205 PNTSY-----PEAT-----QRTT-----RPDLPEQAR----- 229
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db 213 PKDNLLDMCPSPATPALGSELQMLPKSRLLTVNVCYSIQDPPSSVNLNNSGKP 272

QY 230 -----RSATWTHSH-----PT-----QSKAQPSSS 250
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db 273 KDHSFPENGSGSFESSDILLRSMWNSQSSLLDVQVRPSEESFEEDCSQSGLSKRLTMSFED 332

QY 251 TVPKTEDROPOLDYQIILGPTSSRLANSGSGIQLMQFLLELLTSDSSNSNCITWEGTNGE 310
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db 333 YIGERSDPEVEGKPF---YIPAAVLAGFLFGSGGIQLMQFLLELLTSLSKSCSOFISWIGDNGE 389

QY 311 EKMTDPEVARRWGERKSKPNMNTDKLSRALRYYYDKNMIMTKVHGKRYAYKF 362
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:

```

```

Db      390 FRLADPDEVARRMGRKRNKPKNMYEKLNSGLGYYDDKNIHHTSGRRYYRF 441

RESULT 7
US-09-092-636-4
; Sequence 4, Application US/09092636A
; Patent No. 6162641
; GENERAL INFORMATION:
; APPLICANT: Goldman, Daniel
; APPLICANT: Sapru, Mohan K.
; TITLE OF INVENTION: Neuregulin Response Element and Uses Therefor
; FILE REFERENCE: UMI-003
; CURRENT APPLICATION NUMBER: US/09/092,636A
; CURRENT FILING DATE: 1998-06-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 470
; TYPE: PRT
; ORGANISM: Rattus sp.
US-09-092-636-4

Query Match          17.2%; Score 421.5; DB 4; Length 470;
Best Local Similarity 31.5%; Fred. No. 8.4e-32;
Matches 112; Conservative 39; Mismatches 84; Indels 121; Gaps 11.

QY    122 ERRVVPADPTLMSIDNHQWLEMWVKFEGLPDVIDLLFQNTDGKELCMTKDDORLTLP 181
       ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     94 QRRIGCPKRPWTLMNDOQCQMLHMATNERSLVNVNLQRFGNGQMCLMGKERFLIELAP 152

QY    182 SYNDILSLHLHYLERGAT-----FTFPNTSVY-----                210
       ||| ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     153 DEVGDLIMELHLOMKKENOEKTEDOYEENSHLNAAVPHWNSTWLTFGVBOALPYGMOPSY 212

QY    211 -----PEAT-----QRLTTRPDLYEDARBSAWT-----SSHPTOSKATQ 246
       ||| ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     213 LKDGLDMCKPPSATPALGSEQLOMLP-----KSRLTVSVNCTCSIQDFPGGNLTL 267

QY    247 PSSSTVPTED-----OR-POLDPYO-----IIG 269
       ||| ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     268 NSSGCKPREHDSPENGCDSESSDSLNSWSQSLLDVQVRPSFEDECSQSLCJSK 327

QY    270 PRSS-----RANP-----GSGQIQLMQFLLELLSDSSNSNCITWEG 306
       ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     328 PMSFKDYIQERSDVEEQCKPIPAVALAGFTGSGPIQLMQFLLELLSDSKSQSFSIWGTG 387

QY    307 TNGCEKMTDPDEVARRMGERKSPMMNYDKLSRALRYYYDKNKIMTKVHGKRYYRF 362
       ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     368 DGMERKLADPDEVARRMGRKRNKPKNMYEKLNSGLGYYDDKNIHHTSGRRYYRF 443

RESULT 8
US-08-469-412A-7
; Sequence 7, Application US/08469412A
; Patent No. 5856125
; GENERAL INFORMATION:
; APPLICANT: Mavrothalassitis, George J.
; APPLICANT: Blair, Donald G.
; APPLICANT: Fisher, Robert J.
; APPLICANT: Beal Jr., Gregory J.
; APPLICANT: Athanasiou, Metropi A.
; APPLICANT: Scouras, Dionysios N.
; TITLE OF INVENTION: The ERF Genetic Locus and Its Products
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

```

1      COMPUTER: IBM PC compatible
2      OPERATING SYSTEM: PC-DOS/MS-DOS
3      SOFTWARE: Patentin Release #1.0, Version #1.30
4      CURRENT APPLICATION DATA:
5          APPLICATION NUMBER: US/08/469,412A
6          FILING DATE: 05-JUN-1995
7          CLASSIFICATION: 435
8      ATTORNEY/AGENT INFORMATION:
9          NAME: Garrett-Wackowski, Eugenia
10         REGISTRATION NUMBER: 37,330
11         REFERENCE/DOCKET NUMBER: 015280-229000
12         TELECOMMUNICATION INFORMATION:
13             TELEPHONE: (415) 576-0200
14             TELEFAX: (415) 576-0300
15     INFORMATION FOR SEQ ID NO. 7:
16         SEQUENCE CHARACTERISTICS:
17             LENGTH: 543 amino acids
18             TYPE: amino acid
19             STRANDEDNESS:
20                 TOPOLOGY: linear
21         MOLECULE TYPE: protein
22         FEATURE:
23             NAME/KEY: Protein
24             LOCATION: 1..543
25             OTHER INFORMATION: /note= "murine erf amino acid sequence
26             OTHER INFORMATION: (first 8 amino acids from first exon not
27             OTHER INFORMATION: included)"
28     US-08-469-412A-7
29
30 Query Match                13.9%; Score 340.5; DB 24; Length 543;
31 Best Local Similarity      42.5%; Pred. No. 5.7e-24;
32 Matches 71; Conservative 19; Mismatches 44; Indels 33; Gaps 3
33
34 QY 276 ANPGSGOIOLMQLLELLSDSSNSNCITWEGTNGEFKMTDPDEVARRMGERSKPMNYD 335
35      ::::|::::|::::|::::|::::|::::|::::|::::|::::|::::|::::|::::|
36 Db 12 SSPGSRQIQLMFLIELLRKEEYGIVAMQGDYGEVIRKDPEVARLIMGVRCRCKQMND 71
37      |||||||||:|-:-|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
38 QY 336 KLSRLRRYYDKNINIKTYHGKRKYAKFDH-----GIAQLQHPHES 378
39      |||||||||:|-:-|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
40 Db 72 KLSRLRRYYNRKLHKTKGRFYFKFNKLVLYNPIDMWGLAGGAVPQSAPVPSSG 131
41      |::::|-----SEVLSPTEDDPPSPACSSSSSLPSFA 167
42
43 RESULT 9
44 US-09-021-715-7
45 ; Sequence 7, Application US/09021715
46 ; Patent No. 6194547
47 ; GENERAL INFORMATION:
48 ; APPLICANT: Mavrothalassitis, George J.
49 ; Blair, Donald G.
50 ; Fisher, Robert J.
51 ; Neal Jr., Gregory J.
52 ; Athanasiou, Meropi A.
53 ; Sgouras, Dionysios N.
54 TITLE OF INVENTION: The ERF Genetic Locus and Its Products
55 NUMBER OF SEQUENCES: 16
56 CORRESPONDENCE ADDRESS:
57 ADDRESSEE: Townsend and Townsend and Crew LLP
58 STREET: Two Embarcadero Center, Eighth Floor
59 CITY: San Francisco
60 STATE: California
61 COUNTRY: USA
62 ZIP: 94111-3834
63
64 COMPUTER READABLE FORM:
65 MEDIUM TYPE: floppy disk
66 COMPUTER: IBM PC compatible
67 OPERATING SYSTEM: PC-DOS/MS-DOS
68 SOFTWARE: Patentin Release #1.0, Version #1.30
69 CURRENT APPLICATION DATA:
70 APPLICATION NUMBER: US/09/021,715

```

RESULT 10 412A-2
 US-08-469-412A-2
 Sequence 2, Application US/08469412A
 Patent No. 5856125
 GENERAL INFORMATION:
 APPLICANT: Mavrothlassitis, George J.
 APPLICANT: Blair, Donald G.
 APPLICANT: Fisher, Robert J.
 APPLICANT: Beal Jr., Gregory J.
 APPLICANT: Athanasiou, Merope A.
 APPLICANT: Sgouras, Dionysios N.
 TITLE OF INVENTION: The ERF Genetic Locus and Its Products
 NUMBER OF SEQUENCES: 16
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Townsend and Townsend and Crew LLP
 STREET: Two Embarcadero Center, Eighth Floor
 CITY: San Francisco
 STATE: California
 COUNTRY: USA
 ZIP: 94111-3834
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/469,412A
 FILING DATE: 05-JUN-1995
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Garrett-Wackowski, Eugenia

RESULT 11
 US-09-021-715-2
 Sequence 2, Application US/09021715
 Patent No. 6194547
 GENERAL INFORMATION:
 APPLICANT: Mavrochaitlis, George J.
 Blair, Donald G.
 Fisher, Robert J.
 Beal Jr., Gregory J.
 Athanasiou, Metropi A.
 Sgouras, Dionysios N.
 TITLE OF INVENTION: The ERF Genetic Locus and Its Products
 NUMBER OF SEQUENCES: 16
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Townsend and Townsend and Crew LLP
 STREET: Two Embarcadero Center, Eighth Floor
 CITY: San Francisco
 STATE: California
 COUNTRY: USA
 ZIP: 94111-3834
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/021,715
 FILING DATE: 10-Feb-1998
 CLASSIFICATION: <Unknown>
 ATTORNEY/AGENT INFORMATION:
 NAME: Garrett-Wackowski, Eugenia
 REGISTRATION NUMBER: 37,330
 REFERENCE/DOCKET NUMBER: 015280-229000
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 576-0200
 TELEFAX: (415) 576-0300
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 548 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 SEQUENCE DESCRIPTION: SEQ ID NO: 2:


```
|: |||::: |: | :||| ||||| ||||| |||||
Db 23 YDSVRRGAWGNMNSGLNKSPPLGAGOTISKTEORRPPDYQILGPTSSRLANPGOI 82
OY 284 QLMOFLELLSDSNSNCITWEGTNGE 310
||| ||||| ||||| ||||| ||||| |||||
Db 83 QLMOFLELLSDSANASCITWEGTNGE 109
```

Search completed: November 9, 2002, 16:55:36
Job time : 23.3751 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 9, 2002, 09:11:56 ; Search time 38.1625 Seconds

(without alignments)
7638.889 Million cell updates/sec

Title: US-09-902-772-1_COPY_645_662

Perfect score: 18
Sequence: 1 ctcagagagagagagcc 18

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estda:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_estl:*
9: gb_estl:*
10: gb_est2:*
11: gb_est3:*
12: gb_est4:*
13: gb_est5:*
14: gb_est6:*
15: em_estfun:*
16: em_estlom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pin:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_oher:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	18	100.0	455	17	AO522081 HS_5194_A
2	17	94.4	312	12	BG661425 KX01F12.Y
3	17	94.4	401	10	BE203178 EST403200
4	17	94.4	488	9	AA660203 00044 MCR
5	17	94.4	636	17	AG042409 Pan crog1
C 6	17	94.4	657	17	BH86289 LB00701a.

7	17	94.4	743	17	BH272744	BH272744	CH230-19N
8	17	94.4	944	12	BE962526	BE962526	601655838
9	16.4	91.1	361	12	BF919622	BF919622	RC6-NT015
10	16.4	91.1	380	9	AL700950	AL700950	DRF2686P
11	16.4	91.1	385	10	BE508198	BE508198	dc07b10. x
12	16.4	91.1	458	13	BM149886	BM149886	TCAAP3010
13	16.4	91.1	478	10	AW0702256	AW0702256	TGSTR274
14	16.4	91.1	490	13	BM149408	BM149408	TCAAP2096
15	16.4	91.1	502	14	BQ331367	BQ331367	PK3-ET027
16	16.4	91.1	503	9	AL701675	AL701675	DRF2686P
17	16.4	91.1	561	10	BE261583	BE261583	601149108
18	16.4	91.1	631	10	AL703365	AL703365	DRF2686P
19	16.4	91.1	635	12	BF228672	BF228672	EST00383
20	16.4	91.1	636	9	AL545574	AL545574	AL545574
21	16.4	91.1	709	13	BI758704	BI758704	603024085
22	16.4	91.1	728	9	AU140295	AU140295	AU140295
23	16.4	91.1	758	12	BG679840	BG679840	602626337
24	16.4	91.1	763	9	AL549331	AL549331	AL549331
25	16.4	91.1	771	13	BI753749	BI753749	603028269
26	16.4	91.1	828	10	BE409835	BE409835	601299529
27	16.4	91.1	829	10	BE407601	BE407601	601299520
28	16.4	91.1	849	10	BE640813	BE640813	CL12.1.G1
29	16.4	91.1	868	9	AL549887	AL549887	AL549887
30	16.4	91.1	905	9	AL552363	AL552363	AL552363
31	16.4	91.1	932	12	BG423797	BG423797	602450728
32	16.4	91.1	970	12	BE779108	BE779108	601464713
33	16.4	91.1	981	9	AL556118	AL556118	AL556118
34	16.4	91.1	1004	17	CNS0584M	CNS0584M	AL350095
35	16.4	91.1	1119	17	B09253	B09253	F11DB-SP6
36	16.4	91.1	1134	12	BG773107	BG773107	602721827
37	16.4	91.1	1193	13	BE963390	BE963390	601657157
38	16.4	91.1	1243	12	BE945890	BE945890	NR3-KT005
39	16.4	91.1	1250	17	FR0054774	FR0054774	AL691053
40	16.4	91.1	1259	12	BE618455	BE618455	601462876
41	16.4	91.1	1259	12	BG477769	BG477769	602521483
42	16.4	91.1	1275	9	AV258535	AV258535	AV258535
43	16.4	91.1	1275	9	AA736629	AA736629	NW53C09.s
44	16.4	91.1	1334	9	AA248965	AA248965	KK6770.se
45	16.4	91.1	1380	17	AQ970497	AQ970497	RRC1-23-3

ALIGNMENTS

RESULT 1
AO522081/c
LOCUS
DEFINITION
HS_5194_A1.D04.T7A RPCI-11 Human Male BAC Library Homo sapiens
genomic clone Plate=770 Col=7 Row=G, DNA sequence.
ACCESSION
AO522081
VERSION
AO522081.1 GI:4769203
KEYWORDS
GSS.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 455)
Mahairas G.G., Wallace J.C., Smith K., Swartzell S., Holzman T.,
Keller A., Shaker R., Furlong J., Young J., Zhao S., Adams M.D. and
Hood L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
JOURNAL
MEDLINE
COMMENT
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong

(pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm) or from Research Genetics (<http://www.htsc.washington.edu>). BAC end Web Server: <http://www.htsc.washington.edu>

Plate: 770 row: G column: 7

Seq primer: 17

Class: BAC ends

High quality sequence stop: 455.

FEATURES
source Location/Qualifiers

1..455

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="RPGT-11 Human Male BAC Library"

/sex="male"

/note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACE3.6 vector at EcoRI sites"

BASE COUNT 102 a 121 c 100 g 126 t 6 others

ORIGIN

Query Match 100.0%; Score 18; DB 17; Length 455;

Best Local Similarity 100.0%; Pred. No. 9.5e+02;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCAGAGAGAGAGACC 18

Db 153 CTCAGAGAGAGAGACC 136

RESULT 2
LOCUS B661425 312 bp mRNA linear EST 21-JUN-2001

DEFINITION Kx01f12.y1 Parastromyoloides trichosuri IL SL1 TOPO v2

Parastromyoloides trichosuri CDNA, mRNA sequence.

ACCESSION B661425.1 GI:13817193

VERSION B661425.1 GI:13817193

KEYWORDS EST.

SOURCE Parastromyoloides trichosuri.

ORGANISM Parastromyoloides trichosuri.

REFERENCE 1 (bases 1 to 312)

AUTHORS McCarter, J., Clifton, S., Chiapelli, B., Pape, D., Martin, J., Wylie, T., Dante, M., Marra, M., Hillier, L., Kucaba, T., Theising, B., Bowers, Y., Gibbons, M., Rittner, E., Bennett, J., Franklin, C., Tsagaris, V., R., Ronko, I., Kennedy, S., Maguire, L., Beck, C., Underwood, K., Steptoe, M., Allen, M., Person, B., Swaller, T., Harvey, N., Schurk, R., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R., and Wilson, R.

The Washington Univ. Nematode EST Project, 1999
Unpublished (1999)
Contact: McCarter JP
The Washington Univ. Nematode EST Project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu

Nematodes provided by Dr. Warwick Grant of AgResearch, New Zealand (warwick.grant@agresearch.co.nz). The library was constructed by Brandi Chiapelli and Dr. James McCarter (bchiapelli@watson.wustl.edu & jmcarter@watson.wustl.edu) at Washington University, St. Louis. DNA Sequencing by: Washington University Genome Sequencing Center, St. Louis.

Seq primer: SL1 primer.

FEATURES
source Location/Qualifiers

1..312

/organism="Parastromyoloides trichosuri"

/db_xref="taxon:131310"

/clone_lib="Parastromyoloides trichosuri IL SL1 TOPO v2"

/dev_stage="Infective larvae"

/lab_host="DH10B"

/note="Vector: PCR11-TOPO; SL1-Oligo(dT) PCR-based library. Parastromyoloides trichosuri CDNA PCR products of size >400 nucleotides containing SL1 on the 5' end and oligo(dT) on the 3' end were non-directionally cloned into PCR11-TOPO (Invitrogen) following the TOPO TA cloning protocol. The CDNA insert can be excised by digestion with EcoRI. The library was constructed by Brandi Chiapelli and Dr. James McCarter at Washington University, St. Louis."

BASE COUNT 58 a 118 c 66 g 70 t

ORIGIN

Query Match 94.4%; Score 17; DB 12; Length 312;

Best Local Similarity 100.0%; Pred. No. 2.1e+03;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TCAGAGAGAGAGACC 18

Db 224 TCAGAGAGAGAGACC 240

RESULT 3
LOCUS BE203178 401 bp mRNA linear EST 07-SEP-2000

DEFINITION EST403200 KVL Medicago truncatula CDNA clone pKV1-4P5, mRNA sequence.

ACCESSION BE203178.1 GI:8746449

VERSION BE203178

KEYWORDS EST.

SOURCE barrel medic.

ORGANISM Medicago truncatula

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae; Medicago.

1 (bases 1 to 401)

AUTHORS Vandenbosch, K., Endre, G., Hur, J., Moore, J., Beremand, P., Ellis, L., Town, C.D., Bowman, C.L., Craven, M.B., Hansen, T.S., Holt, I.E., and Fraser, C.M.

ESTs from roots of Medicago truncatula 24 hours after inoculation with Sinorhizobium meliloti

Unpublished (1999)

Contact: Vandenbosch K

Department of Plant Biology

University of Minnesota

220 Biosci Center, 1445 Gortner Ave, St. Paul, MN 55108, USA

Tel: 612 624 2755

Fax: 612 625 1738

Email: kvandenb@cbbs.umn.edu

Texas A&M University: T262950e

TIGR sequence name: MTIAK877K

More information is available at:

<http://chrysis.tamu.edu/medicago>

Seq primer: SKmod (CTA GAA CTA GCG GAT CC).

Location/Qualifiers

FEATURES
source

1..401

/organism="Medicago truncatula"

/cultivar="genotype A17"

/db_xref="taxon:3880"

/clone_lib="KV1-4P5"

/clone_lib="KV1"

/tissue_type="Seedling roots"

/dev_stage="24 hours post-inoculation with Sinorhizobium meliloti"

/lab_host="E. coli strain XL0R"

/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; CDNA was prepared from polyA+ enriched RNA. The CDNA was directionally ligated into the unzip XR vector from

stratagene and packaged using Gigapack III Gold packaging

extracts. Plasmids containing CDNA inserts were excised

from the recombinant lambda-Zap phage using Ex-assist

helper phage and propagated in XL0R cells."

BASE COUNT 114 a 86 c 128 g 73 t
ORIGIN
Query Match 94.4%; Score 17; DB 10; Length 401;
Best Local Similarity 100.0%; Pred. No. 2.4e+03;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 TCAGAGAGAGAGGCC 18
|||||
Db 287 TCAGAGAGAGAGGCC 303
|||||
RESULT 4
AA660203 488 bp mRNA linear EST 08-MAR-2000
DEFINITION 00044 MREHE Medicago truncatula cDNA similar to phosphoenolpyruvate
carboxylase, mRNA sequence.
ACCESSION AA660203
VERSION AA660203.1 GI:2604247
KEYWORDS EST.
SOURCE Medicago medic.
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.
1 (bases 1 to 488)
Coviltz,P.A., Smith,L.S. and Long,S.R.
Expressed sequence tags from a root-hair-enriched medicago
truncatula cDNA library
Plant Physiol. 117 (4), 1325-1332 (1998)
CONTACT: Long SR
Institute of Biological Sciences and Howard Hughes Medical
Institute
Stanford University
Gilbert Biology, Stanford, CA 94305-5020, USA
Tel: 650 723 3232
Fax: 650 725 8309
Email: fa.srl@forsthe.stanford.edu
Seq primer: T3.
FEATURES
Location/Qualifiers
1..488
/organism="Medicago truncatula"
/cultivar="Jemalong"
/db_xref="taxon:3880"
/clone_lib="MREHE"
/tissue_type="Root hairs & tips"
/dev_stage="2-3 day old seedlings"
/note="Organ: Root; Vector: pBK-CMV; Site_1: EcorI;
Site_2: XhoI; cDNA was synthesized from a pooled mRNA prep
from elongating root hairs (30% w/w) and 2-3cm root tips
(70% w/w). XhoI-oligo-dT linker-primer and EcorI
adaptors were used. cDNAs was cloned unidirectionally
into lambda ZAP Express (Stratagene), amplified, and
mass-excised into pBK-CMV vector plasmids. More
information is available at <http://bio-SR8.stanford.edu>."
BASE COUNT 147 a 103 c 142 g 84 t 12 others
ORIGIN
Query Match 94.4%; Score 17; DB 9; Length 488;
Best Local Similarity 100.0%; Pred. No. 2.6e+03;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 TCAGAGAGAGAGGCC 18
|||||
Db 361 TCAGAGAGAGAGGCC 377
|||||
RESULT 5
AG042409 636 bp DNA linear GSS 01-NOV-2001
LOCUS AG042409/c
DEFINITION Pan troglodytes DNA, clone: PTB-020G15.F, genomic survey sequence.
ACCESSION AG042409

VERSION AG042409.1 GI:16571134
KEYWORDS GSS.
SOURCE Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
BAC Library clone:PTB-020G15.F.
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Pan.
REFERENCE
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE BAC end sequences of library PTB
JOURNAL
REFERENCE
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suenhiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:chimpbes@gsc.riken.go.jp, URL:<http://hgp.gsc.riken.go.jp/>,
Tel:81-45-503-9111, Fax:81-45-503-9170)
Clones are derived from the chimpanzee BAC library PTB this BAC end
was generated during the R&D process and may have higher chance of
clone tracking errors.
PRIMERS
Sequencing: -21M13
LIBRARY
Vector : pKS145
R.Site 1 : SacI
R.Site 2 : SacI
FEATURES
source
Location/Qualifiers
1..636
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone_lib="PTB-020G15.F"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT 151 a 173 c 160 g 152 t
ORIGIN
Query Match 94.4%; Score 17; DB 17; Length 636;
Best Local Similarity 100.0%; Pred. No. 3e+03;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 TCAGAGAGAGAGGCC 18
|||||
Db 336 TCAGAGAGAGAGGCC 320
|||||
RESULT 6
BH886289 657 bp DNA linear GSS 07-AUG-2002
LOCUS BH886289/c
DEFINITION LB00701a.d.T7.1 Leishmania major Friedlin BAC Library Leishmania
major genomic clone LB00701a, DNA sequence.
ACCESSION BH886289
VERSION BH886289.1 GI:22130684
KEYWORDS GSS.
SOURCE Leishmania major.
ORGANISM Leishmania major
Eukaryota; Euzoenozoa; Kinetoplastida; Trypanosomatidae;
Leishmania.
REFERENCE
AUTHORS Myler,P.J., Vogt,C., Munden,H., Robertson,L., Sisk,E., Fazelinia,G.,
Aggarwal,G., Nelson,S., Seyler,A., Worthley,E., Stuart,K. and
Ragland,M.
TITLE Leishmania major Friedlin BAC End Sequences
JOURNAL Unpublished (2002)
COMMENT Other GSSs: LB00701a.d.sp6.1
Seattle Biomedical Research Institute
4 Nickerson Street, Seattle, WA 98109-1651, USA
Tel: 206 284-8846

Fax: 206 284-0313
Email: mylerj@sdri.org
Seq primer: T7

Class: BAC ends.

FEATURES
source

Location/Qualifiers
1..657
/organism="Leishmania major"
/strain="Friedlin"
/db_xref="taxon:5664"
/clone_lib="LB00701a"
/clone_lib="Leishmania major Friedlin BAC Library"
/lab_host="E. coli GeneHogs + Trifa"
/note="Vector: pCG270; Site_1: HindIII; Genomic DNA from Leishmania major Friedlin in agarose blocks was partially digested with HindIII, size selected, and ligated with HindIII-digested pCG270 vector DNA. 10368 clones were picked and arrayed in 384- and 96-well plates. Library construction and arraying was carried out by ResGen Corporation and clones and filters are available from them"

BASE COUNT 101 a 228 c 172 g 156 t
ORIGIN

Query Match 94.4%; Score 17; DB 17; Length 657;
Best Local Similarity 100.0%; Pred. No. 3e+03;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CTCAGAGAGAGAGAGC 17
|||||
Db 600 CTCAGAGAGAGAGAGC 584

RESULT 7
BH272744
LOCUS BH272744
DEFINITION CH230-19N18. TV CHORI-230 Segment 1 Rattus norvegicus genomic clone
CH230-19N18, DNA sequence.
ACCESSION BH272744
VERSION BH272744.1 GI:17185146
KEYWORDS GSS.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 743)
Zhao, S., Shetty, D., Shatsman, S., Tsegaye, G., Geer, K., Shvartsbeyn, A., Gebregoryis, E., Overton, L., Russell, D., Chen, D., Riggs, F., de Jong, P. and Fraser, C.M.
Rat BAC End Sequences from Library CHORI-230 EcorI segment
Unpublished (1999)
Other_GSSs: CH230-19N18.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@ligr.org
Clones are derived from the rat BAC library CHORI-230
(http://www.chori.org/bacpac/rat230.htm). For BAC library availability, please contact Pieter de Jong (pje@onjemail.cho.org).
Clones may be purchased from BACPAC Resources
(http://www.chori.org/bacpac/orering_information.htm). BAC end
page: http://www.ligr.org/ldb/bac_ends/rat/bac_end_intro.html
Plate: 19 row: N column: 18
Seq primer: T7
Class: BAC ends.

FEATURES
source

Location/Qualifiers
1..743
/organism="Rattus norvegicus"
/strain="BN/SsNHsd/MC/M"
/db_xref="taxon:10116"

/clone="CH230-19N18"
/clone_lib="CHORI-230 Segment 1"
/sex="Female"
/cell_type="Brain"

/note="Vector: pTARBAC2.1; Site_1: EcorI; Site_2: EcorI;
CHORI-230 Rat (BN/SsNHsd/MC/M) BAC library produced by
Pieter de Jong"

BASE COUNT 226 a 165 c 158 g 194 t
ORIGIN

Query Match 94.4%; Score 17; DB 17; Length 743;
Best Local Similarity 100.0%; Pred. No. 3.2e+03;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 TCAGAGAGAGAGAGGCC 18
|||||
Db 570 TCAGAGAGAGAGAGGCC 586

RESULT 8
BE962526
LOCUS BE962526
DEFINITION BE962526 944 bp mRNA linear EST 14-DEC-2000
601655838R1 NTH_MGC_66 Homo sapiens CDNA clone IMAGE:3855872 3',
mRNA sequence.
ACCESSION BE962526
VERSION BE962526.2 GI:11765511
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 944)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
On Oct 3, 2000 this sequence version replaced gi:10573231.
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: DCTD/DRP
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LDCM570 row: k column: 09
High quality sequence stop: 110.

FEATURES
source

Location/Qualifiers
1..944
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="IMAGE:3855872"
/clone_lib="NTH_MGC_66"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: ovary; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.8 kb. Library constructed by Life
Technologies."

BASE COUNT 188 a 240 c 254 g 262 t
ORIGIN

Query Match 94.4%; Score 17; DB 12; Length 944;
Best Local Similarity 100.0%; Pred. No. 3.6e+03;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 TCAGAGAGAGAGAGGCC 18
|||||
Db 727 TCAGAGAGAGAGAGGCC 743

RESULT 9
BF919622

LOCUS BF919622 361 bp mRNA linear EST 19-JAN-2001
DEFINITION RC6-NT0157-011100-011-E02 NT0157 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF919622
VERSION BF919622.1 GI:12315510
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 361)
Dias Neto, E., Garcia Correia, R., Verjovsky-Almeida, S., Briones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, R.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,
Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=RC6&t2=RC6-NT0157-
011100-011-E02&t3=2000-11-01&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 8
High quality sequence stop: 69.
Location/Qualifiers
source
1..361
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NT0157"
/dev_stage="Adult"
/note="Organ: nervous_tumor; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 62 a 105 c 117 g 77 t
ORIGIN
Query Match 91.1%; Score 16.4; DB 12; Length 361;
Best Local Similarity 94.4%; Pred. No. 4.1e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 CTCAGAGAGAGAGAGCC 18
|||||
Db 300 CTCAGAGAGAGAGAGCC 317
RESULT 10 380 bp mRNA linear EST 21-MAR-2002
AL700950
LOCUS DKFZp686P0711.1 686 (synonym: hlcc3) Homo sapiens cDNA clone
DEFINITION DKFZp686P0711.1 5', mRNA sequence.
ACCESSION AL700950
VERSION AL700950.1 GI:19621483
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 380)

AUTHORS Bloeker, H., Boecker, M., Brandt, P., Mewes, W., Well, B. and Wiemann
, S.
TITLE EST (Bloeker, H., Boecker, M., Brandt, P., Mewes, H.W., Well, B. and
Wiemann, S.)
JOURNAL Unpublished (1999)
COMMENT Contact: Bloecker H
MIPS
Am Klopferstritz 18a D-82152 Martinsried, Germany
This is the 5' sequence of the clone insert
clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ), Email s.wiemann@dkfz-heidelberg.de,
sequenced by GAP (National Research Centre for Biotechnology Ltd.,
Braunschweig/Germany) within the cDNA sequencing consortium of the
German Genome Project.
No s1 sequence available.
This clone (DKFZp686P0711) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcentzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
Location/Qualifiers
source
1..380
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFZp686P0711"
/clone_lib="686 (synonym: hlcc3)"
/issue_type="human skeletal muscle"
/tissue_type="human skeletal muscle"
/dev_stage="adult"
/lab_host="DH10B"
/note="Vector: pT7Blue2; Site_1: SfiIA; Site_2: SfiIB;
cDNA-collection"
BASE COUNT 76 a 115 c 111 g 78 t
ORIGIN
Query Match 91.1%; Score 16.4; DB 9; Length 380;
Best Local Similarity 94.4%; Pred. No. 4.2e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 CTCAGAGAGAGAGAGCC 18
|||||
Db 104 CTCAGAGAGAGAGAGCC 121
RESULT 11 385 bp mRNA linear EST 07-AUG-2000
BE508198
LOCUS dc07b10.x1 NICHD XGC L11 Xenopus laevis cDNA clone IMAGE:3396379 3'
DEFINITION similar to SW:CO3_XENLA P23667 COMPLEMENT C3 ;, mRNA sequence.
ACCESSION BE508198
VERSION BE508198.1 GI:9727973
KEYWORDS EST.
SOURCE African clawed frog.
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
Xenopodidae; Xenopus.
1 (bases 1 to 385)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Other_ESTs: dc07b10.y1
Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Martha Rebert, Steven L. Klein, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Washington University Genome Sequencing Center
through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov
Trace considered overall poor quality
Seq primer: -40UP from Gibco
High quality sequence stop: 1.
Location/Qualifiers
source
1..385

```

/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone="IMAGE:3396379"
/clone_lib="NICHD XCG L11"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: liver; Vector: pCMV-Sport6; Site_1: NotI; Site_2: SalI; Cloned unidirectional. Primer: Oligo dr. Average insert size 1.4 Kb. Constructed by Life Technologies. Note: This is a Xenopus Gene Collection (XCG) library."
BASE COUNT      94 a      91 c      136 g      64 t
ORIGIN
Query Match      91.1%; Score 16.4; DB 10; Length 385;
Best Local Similarity 94.4%; Pred. No. 4.2e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY      1 CTCAGAGAGAGAGAGGCC 18
||| |||||||||||||||
Db      333 CTCGAGAGAGAGAGAGGCC 350

```

RESULT	12
BMI49896	
LOCUS	BMI49896
DEFINITION	BMI49896 458 bp mRNA linear EST 30-NOV-2001 TCAP2P10590 Pediatric acute myelogenous leukemia cell (FAB M1) Bay1or-HSC project-TCA Homo sapiens CDNA clone TCAP1059, mRNA sequence.
ACCESSION	BMI49896
VERSION	BMI49896
KEYWORDS	BMI49896.1 GI:17171920
SOURCE	EST. human.

ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
Homo sapiens					
Eukaryotes; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.					
1 (bases 1 to 458)					
Wei, Y., Tsang, Y. T. M., Mei, G., Ku, J. M., Ali-Osman, F. R., Jr.,					
Guarnierine, P. H., Muzny, D., Boucek, J. J., Gibbs, R. A., and Margolin, J. F.					
Pediatric Leukemia cDNA Sequencing Project (2001)					
Unpublished (2001)					
Contact: Dr. Judith F. Margolin					

Texas Children's Cancer Center and Human Genome Sequencing Center
 at Baylor College of Medicine
 1102 Bates, MC3-3330 Houston, TX 77030, USA
 Tel.: 832-824-4536
 Fax: 832-825-4038
 Email: clones@ccc.org
 Seq primer: M13 primer.
 Location/Qualifiers
 source
 1.458

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="TCAP1059"
/clone_lib="pediatric acute myelogenous leukemia cell (FDB
M1) Baylor-HSC project-TCMA"
/sex="male"
/tissue_type="leukopheresis"
/cell_type="myeloid cell"
/dev_stage="pediatric 6 years"
/lab_host="DH10B"
/notice=Vector: lambda PSB, site_1: BamHI; site_2: EcoRI;
first strand cDNA was primed with an anchored
XhoI-oligo(dT) primer [5'GGAGACTCGAGCGCCGACGAGAG(7')VN
3'; V=A,C,G; N=A,C,G,T] and then dg tailed. Second strand
was primed with a BamHI-dC primer
[5'AGAGAGCTCGAGATCCGCGGAGAAATATATAT(C) 3'].
Double-stranded cDNA was then digested with BamHI and XhoII
and directionally cloned into the BamHI and SalI sites of
a lambda psb vector. Library went through one round of
normalization. Library was constructed by Wei Yu at RIKEN
of Japan (Carnegie P, Westover A, Nishiyama Y, Ohsumi T,
Itoh M, Nagaoaka S, Sasaki M, Okazaki Y, Muramatsu M,
Schneider C, Hayashizaki Y, High efficiency selection of

```

```

full-length cDNA by improved biotinylated cap trapper..
DNA Res 4: 1, 61-6, Feb 28, 1997")
BASE COUNT      77 a      146 c      144 g      91 t
ORIGIN
Query Match      91.1%; Score 16.4; DB 13; Length 458;
Best Local Similarity 94.4%; Pred. No. 4.6e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY      1 CTCACAGAGAGAGAGGCC 18
        ||||||| |||||||||
Db      369 CTCACAGGAGAGAGAGGCC 386

```

RESULT	13
LOCUS	AW702256/c
DEFINITION	AW702256 478 bp mRNA linear EST 22-MAY-2000 TGSTGZ74F03.y1 TGrr*-Tachyzoite cDNA Toxoplasma gondii CDNA clone
ACCESSION	TGSTEZ74F03.y1 5', mRNA sequence.
VERSION	AW702256
KEYWORDS	AW702256.1 GI:7586409
SOURCE	EST.
ORGANISM	Toxoplasma gondii. Toxoplasma gondii. Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Saccocystidae; Toxoplasma. 1 (bases 1 to 478) Hell,A., Manger,I., Marra,M., Sidley,L.D., Ajioke,J.A., Aslett,M.A., Dierich,N., Dubugue,T., Hillier,L., Kucaba,T., Wan,K.L., Waterton,R.H. and Boothroyd,J. Mashu-Merck-stanford-NIH Toxoplasma EST project Unpublished (1996) Contact: Marra M
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
COMMENT	

WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: toxoe@watson.wustl.edu
Contact David Sibley (toxoe@borcim.wustl.edu) for further
information relating to organism, libraries, or clone availability.
Seq primer: -40RP from GIBCO
High quality sequence stop: 319.
Location/Qualifiers
1. .478

```

/organism="Toxoplasma gondii"
/strain="RH (Type I)"
/db_xref="taxon:5811"
/clone="TgEST274f03.y1"
/clone_id="TgRH+Tachyzoite cDNA"
/dev_stage="Tachyzoite"
/lab_host="SOLR cells"
/note="Vector: pluescript SK-; Site_1: EcoRI; Site_2:
XhoI; cDNA was synthesized from poly mRNA using an
oligo-dT primer containing an XhoI site. Following second
strand synthesis, EcoRI adapters were ligated to the cDNA
and products were size-selected on sephacryl S500. The
cDNAs were ligated to EcoRI/ XhoI prepared lambda ZapII
(Stratagene). Clones were converted to phagemids by mass
excision using EXAssist helper phage and E.coli SOLR cells
(Stratagene). Insert sizes range from 0.3-3.0 kb. The
library may contain a small percentage of host or
bacterial contaminants. NOTE: Many clones lack the 5'
EcoRI site but contain the upstream polylinker sites."
BASE COUNT      69 a      166 c      112 g      131 t
ORIGIN
Query Match      91.1%; Score 16.4; DB 10; Length 478;
Best Local Similarity 94.4%; Pred. No. 4.7e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
1 CTCGAGAGAGAGAGAGACC 18
|||||

```

Db 443 CTCAGAGGAGAGGAGAC 426

RESULT 14
BML149408

LOCUS
DEFINITION BML149408 490 bp mRNA linear EST 30-NOV-2001
TCMAA209618 Pediatric acute myelogenous leukemia cell (FAB M1)
Baylor-HGSC project-TCMA Homo sapiens CDNA clone TCMAAP618, mRNA
sequence.
BML149408
ACCESSION BML149408.1 GI:17170993
VERSION EST.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 490)
AUTHORS Mei,Y., Tsang,Y.T.M., Mei,G., Ku,J.M., Ali-Osman,F.R. Jr.,
Gunnarathne,P.H., Muzny,D., Bouck,J., Gibbs,R.A. and Margolin,J.F.
Pediatric Leukemia CDNA Sequencing Project (2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Dr. Judith F. Margolin
Texas Children's Cancer Center and Human Genome Sequencing Center
at Baylor College of Medicine
1102 Bates, MC3-3320 Houston, TX 77030, USA
Tel: 832-824-4536
Fax: 832-825-4038
Email: clones@ccc.org
Seq primer: M13 primer.

FEATURES
source
1..490
location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="TCMAAP618"
/clone_1lb="Pediatric acute myelogenous leukemia cell (FAB
M1) Baylor-HGSC project-TCMA"
/sex="male"
/tissue_type="leukopheresis"
/cell_type="myeloid cell"
/dev_stage="pediatric 6 years"
/lab_host="DH10B"
/note="Vector: lambda PSB. Site_1: BamHI; Site_2: EcoRI;
First strand cDNA was primed with an anchored
XhoI-oligo(dt) primer [5'GAGAGACTCGAGCGCCGCGAGAGAG(T)VN
3'; V-A,C,G; N-A,C,G,T] and then dg tailed. Second strand
was primed with a BamHI-dc primer
[5'AGAGAGCTCGAGTCGCGCGCCGCAATTAATAT(C) 3'].
Double-stranded cDNA was then digested with BamHI and XhoI
and directionally cloned into the BamHI and SalI sites of
lambda PSB vector. Library was constructed through one round of
normalization. Library was constructed by Wei Yu at RIKEN
of Japan (Garninci P, Westover A, Nishiyama Y, Ohsumi T,
Itoh M, Nagaoaka S, Sasaki, Okazaki Y, Muramatsu M,
Schneider C, Hayashizaki Y, High efficiency selection of
full-length cDNA by improved biotinylated cap trapper.,
DNA Res 4: 1, 61-6, Feb 28, 1997)."

BASE COUNT 77 a 156 c 165 g 92 t

ORIGIN

Query Match 91.1%; Score 16.4; DB 13; Length 490;
Best Local Similarity 94.4%; Pred. No. 4.7e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTCAGAGAGAGAGAGCC 18
||||| |||||||

Db 449 CTCAGAGGAGAGAGAGCC 466

RESULT 15
BO331367

LOCUS
DEFINITION BO331367 502 bp mRNA linear EST 17-MAY-2002
PM3-ET0272-300301-005-g12 ET0272 Homo sapiens CDNA, mRNA sequence.
ACCESSION BO331367

VERSION BO331367.1 GI:20972466
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 502)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bata,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
M.J., Soares,F., Brentani,R.R., Reis,J.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663

TITLE
JOURNAL
MEDLINE
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-27049922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PM3&cl=PM3-ET0272-
300301-005-g12&cl=2001-03-30&cl=1)
Seq primer: puc 18 forward
High quality sequence stop: 10.

FEATURES
source
1..502
location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1lb="ET0272"
/dev_stage="Adult"
/note="Organ: lung_tumor; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

BASE COUNT 118 a 119 c 147 g 118 t

ORIGIN

Query Match 91.1%; Score 16.4; DB 14; Length 502;
Best Local Similarity 94.4%; Pred. No. 4.8e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTCAGAGAGAGAGAGCC 18
||||| |||||||

Db 419 CTCAGAGGAGAGAGAGCC 436

Search completed: November 9, 2002, 12:27:52
Job time : 42.1625 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 9, 2002, 15:36:06 ; Search time 53.4015 Seconds

(without alignments)
1125.362 Million cell updates/sec

Title: us-09-902-772-2

Perfect score: 2444
Sequence: 1 MASTIKKALSVSEDSLF.....IIPNRLPAHMPHSILGTY 451

Scoring table: BLOSUM62
Gapop 10.0 , Gapept 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

```

1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2444	100.0	451	20	AAV01520 Chicken C-11 prote
2	2262.5	92.6	478	20	AAV01521 Human C-eryg prot
3	2182	89.3	462	23	AAU75313 Human Ery transcr
4	1624.5	66.5	452	14	AAK4556 Human HUM-FLI-1 ge
5	1599	65.4	451	22	AAU90792 Human shear stress
6	948	38.8	188	22	ABB29050 Peptide #1701 enco
7	948	38.8	188	22	ABB34212 Peptide #1718 enco
8	948	38.8	188	22	ABH19649 Protein #1648 enco
9	948	38.8	188	22	AAH55005 Human brain expres
10	948	38.8	188	22	AAH67391 Human bone marrow

11	948	38.8	188	22	AAH15221 Peptide #1655 enco
12	948	38.8	188	22	AAH27682 Peptide #1719 enco
13	948	38.8	188	22	AAH02964 Peptide #1646 enco
14	948	38.8	188	23	ABG37017 Human peptide enco
15	794	32.5	144	22	ABH1070 Human peptide enco
16	794	32.5	144	22	ABH36259 Peptide #3721 enco
17	794	32.5	144	22	ABH21631 Peptide #3765 enco
18	794	32.5	144	22	AAH57033 Protein #3630 enco
19	794	32.5	144	22	AAH69423 Human bone marrow
20	794	32.5	144	22	AAH29754 Peptide #3791 enco
21	794	32.5	144	22	AAH04949 Peptide #3631 enco
22	599	24.5	475	22	ABH59502 Drosophila melanog
23	562.5	23.0	184	22	ABH82865 Drosophila melanog
24	525	21.5	134	21	AAH01336 Human secreted pro
25	507	20.7	340	22	AAU08495 Rattus norvegicus
26	507	20.7	340	22	AAO21529 Rat lamda73 prote
27	461	18.9	454	14	AAH3365 Sequence of purine
28	454.5	18.6	464	22	ABH62285 Drosophila melanog
29	453.5	18.6	120	22	AAH87650 Bovine mammary tis
30	443.5	18.1	85	23	AAU075320 Human ETS domain c
31	429	17.6	469	21	AAH98040 Human Ets-2. Homo
32	429	17.6	469	21	AAH90689 Human Ets-2 transc
33	421.5	17.2	470	20	AAH78389 Human colon cancer
34	379	15.5	718	22	ABH6061 Rat Ets-2 protein.
35	344.5	14.1	543	18	AAH07702 Drosophila melanog
36	340.5	13.9	543	18	AAH07702 Human prostate can
37	340.5	13.9	548	18	AAH07702 Mouse ETS2 repress
38	340.5	13.9	548	18	AAH07700 Human ETS2 repress
39	327	13.4	84	23	AAU75319 Human Ery transcr
40	325.5	13.3	613	22	ABH62766 Drosophila melanog
41	305	12.5	732	22	ABH59675 Drosophila melanog
42	303	12.4	325	22	AAH85610 Mouse ETS suppress
43	301.5	12.3	109	21	AAH44557 EMS/HUM-FLI-1 fus1
44	300.5	12.3	203	14	AAH54264 Human pancreatic c
45	300.5	12.3	203	23	ABH41242 Human ovarian anti

ALIGNMENTS

```

RESULT 1
ID   AAV01520 standard; Protein; 451 AA.
XX
AC   AAV01520;
XX
DT   23-JUN-1999 (first entry)
XX
DE   Chicken C-11 protein.
XX
XX   Chicken: C-11 protein; cell calcification inhibiting activity;
KW   cell calcification inhibiting agent; c-eryg protein; arthritis deformans;
KW   ossification; spinal column ligament.
XX
OS   Gallus sp.
XX
PN   JP11075871-A.
XX
PD   23-MAR-1999.
XX
PE   29-MAY-1998; 98UP-0166076.
XX
PR   20-JUN-1997; 97US-0050297.
XX
PR   18-JUN-1997; 97US-0878177.
XX
PA   (CHUS ) CHUGAI PHARM CO LTD.
XX   (UYPE-) UNIV PENNSYLVANIA.
XX
DR   WPI; 1999-257708/22.
XX
DR   N-PDB; AAH26551.
XX
PT   An active protein for inhibiting cell calcification - useful for
PT   measuring the calcification of a cell, for diagnosing arthritis

```

PT deformans or ossification of spinal column ligament
XX
PS Claim 1; Page 8; 15pp; Japanese.
XX
CC The present sequence represents a chicken C-11 protein which has cell
CC calcification inhibiting activity. The specification also describes
CC a cell calcification inhibiting agent containing c-erg protein
CC (AA01521). The proteins are used for measuring the calcification of a
CC cell, for diagnosing arthritis deformans or ossification of spinal column
CC ligament.
XX
SQ Sequence 451 AA;

Query Match 100.0%; Score 2444; DB 20; Length 451;
Best Local Similarity 100.0%; Pred. No. 1.2e-211;
Matches 451; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MASTIKALSVSEDSLEFCAGSPHLAKTEMTASSSEYGGTSKMSPRVPOQDWLSOP 60
DB 1 MASTIKALSVSEDSLEFCAGSPHLAKTEMTASSSEYGGTSKMSPRVPOQDWLSOP 60
OY 61 PARVTIKMECNPNQVNGSRNSPDDCSVAKGKVVSSSDVNGMNYGSYMEKHIPPNNMTT 120
DB 61 PARVTIKMECNPNQVNGSRNSPDDCSVAKGKVVSSSDVNGMNYGSYMEKHIPPNNMTT 120
OY 121 NERRVIVPADPTLMSTDHVRQWLEMAVKEGGLPDVDILLFQNDGKELCKMTDDFQRLT 180
DB 121 NERRVIVPADPTLMSTDHVRQWLEMAVKEGGLPDVDILLFQNDGKELCKMTDDFQRLT 180
OY 181 PSYNADILSHLHYLERGATFEPNTSVYPEATORITTRPDLRYOARRSAMTSHHP 240
DB 181 PSYNADILSHLHYLERGATFEPNTSVYPEATORITTRPDLRYOARRSAMTSHHP 240
OY 241 OSKATOPSSSTVKTEDQRPOLDPYOILGPTSSRLANPGSGOIQLOMFLLELSDSNSN 300
DB 241 OSKATOPSSSTVKTEDQRPOLDPYOILGPTSSRLANPGSGOIQLOMFLLELSDSNSN 300
OY 301 CTTWSTNGEFPKMTDPEVARHNGERKSKPNMNYDKLSALRYDYDKNTKVGKRYAV 360
DB 301 CTTWSTNGEFPKMTDPEVARHNGERKSKPNMNYDKLSALRYDYDKNTKVGKRYAV 360
OY 361 KDFHGIQAOLPHPPRESSMYKPSDLPYMSYHAHPQKMFAPHPALPYTSSSFFAA 420
DB 361 KDFHGIQAOLPHPPRESSMYKPSDLPYMSYHAHPQKMFAPHPALPYTSSSFFAA 420
OY 421 PNPYNSPTGGIYPTNRLPAAHMPSHLGTY 451
DB 421 PNPYNSPTGGIYPTNRLPAAHMPSHLGTY 451

RESULT 2
AA01521
ID AA01521 standard; Protein; 478 AA.
XX
AC AA01521;
XX
DT 14-JUN-1999 (first entry)
XX
DE Chicken c-erg protein.
XX
KW Chicken; C-11 protein; cell calcification inhibiting activity;
KW cell calcification inhibiting agent; c-erg protein; arthritis deformans;
KW ossification; spinal column ligament.
XX
OS Gallus sp.
XX
PN JP11075871-A.
XX
PD 23-MAR-1999.
XX
PF 29-MAY-1998; 98JP-0166076.
XX
PR 20-JUN-1997; 97US-0050297.

PR 18-JUN-1997; 97US-0878177.
XX
XX (CHUS) CHUGAI PHARM CO LTD.
XX (UYPE-) UNIV PENNSYLVANIA.
PA
XX
DR WPI: 1999-257708/22.
DR N-PSDB; AAX26552.
XX
XX
PS Claim 5; Page 9-10; 15pp; Japanese.
XX
CC The present sequence represents chicken c-erg protein. The specification
CC also describes a chicken C-11 protein (AA01520) which has cell
CC calcification inhibiting activity and a cell calcification inhibiting
CC agent containing c-erg protein. The proteins are used for measuring the
CC calcification of a cell, for diagnosing arthritis deformans or
CC ossification of spinal column ligament.
XX
SQ Sequence 478 AA;

Query Match 92.6%; Score 2262.5; DB 20; Length 478;
Best Local Similarity 86.3%; Pred. No. 3.1e-195;
Matches 430; Conservative 0; Mismatches 1; Indels 67; Gaps 3;

OY 1 MASTIKALSVSEDSLEFCAGSPHLAKTEMTASSSEYGGTSKMSPRVPOQDWLSOP 60
DB 1 MASTIKALSVSEDSLEFCAGSPHLAKTEMTASSSEYGGTSKMSPRVPOQDWLSOP 60
OY 61 PARVTIKMECNPNQVNGSRNSPDDCSVAKGKVVSSSDVNGMNYGSYMEKHIPPNNMTT 120
DB 61 PARVTIKMECNPNQVNGSRNSPDDCSVAKGKVVSSSDVNGMNYGSYMEKHIPPNNMTT 120
OY 61 PARVTIKMECNPNQVNGSRNSPDDCSVAKGKVVSSSDVNGMNYGSYMEKHIPPNNMTT 120
DB 61 PARVTIKMECNPNQVNGSRNSPDDCSVAKGKVVSSSDVNGMNYGSYMEKHIPPNNMTT 120
OY 121 NERRVIVPADPTLMSTDHVRQWLEMAVKEGGLPDVDILLFQNDGKELCKMTDDFQRLT 180
DB 121 NERRVIVPADPTLMSTDHVRQWLEMAVKEGGLPDVDILLFQNDGKELCKMTDDFQRLT 180
OY 181 PSYNADILSHLHYLER-----GATFIPNTSVYPEA 213
DB 181 PSYNADILSHLHYLERETPLPHLTSDVDKALONSPRLMHARNTGATFIPNTSVYPEA 240
OY 214 TORITTRPDLRYOARRSAMTSHHPQSKATOPSSSTVKTEDQRPOLDPYOILGPTSS 273
DB 214 TORITTRPDLRYOARRSAMTSHHPQSKATOPSSSTVKTEDQRPOLDPYOILGPTSS 273
OY 274 RLANPGSGOIQLOMFLLELSDSNSNCITWSTNGEFPKMTDPEVARHNGERKSKPNMN 333
DB 274 RLANPGSGOIQLOMFLLELSDSNSNCITWSTNGEFPKMTDPEVARHNGERKSKPNMN 333
OY 301 RLANPGSGOIQLOMFLLELSDSNSNCITWSTNGEFPKMTDPEVARHNGERKSKPNMN 360
DB 301 RLANPGSGOIQLOMFLLELSDSNSNCITWSTNGEFPKMTDPEVARHNGERKSKPNMN 360
OY 334 YDKLSALRYDYDKNTKVGKRYAVKDFHGIQAOLPHPPRESSMYKPSDLPYMSY 393
DB 334 YDKLSALRYDYDKNTKVGKRYAVKDFHGIQAOLPHPPRESSMYKPSDLPYMSY 393
OY 361 YDKLSALRYDYDKNTKVGKRYAVKDFHGIQAOLPHPPRESSMYKPSDLPYMSY 400
DB 361 YDKLSALRYDYDKNTKVGKRYAVKDFHGIQAOLPHPPRESSMYKPSDLPYMSY 400
OY 394 -----HAHPQKMFVAPHPALPYTSSSFFAAPNPYNSPTGGIY 433
DB 394 -----HAHPQKMFVAPHPALPYTSSSFFAAPNPYNSPTGGIY 433
OY 401 HGKRYAVKDFHGIQAOLPHPPRESSMYKPSDLPYMSY 460
DB 401 HGKRYAVKDFHGIQAOLPHPPRESSMYKPSDLPYMSY 460
OY 434 PNTRLPAAHMPSHLGTY 451
DB 434 PNTRLPAAHMPSHLGTY 451
OY 461 PNTRLPAAHMPSHLGTY 478
DB 461 PNTRLPAAHMPSHLGTY 478

RESULT 3
AA075313
ID AA075313 standard; Protein; 462 AA.
XX
AC AA075313;
XX
DT 23-APR-2002 (first entry)
XX
DE Human Erg transcription factor.

XX Human: Ery: transcription factor; tumour necrosis factor- α ;
KW TNF- α ; interleukin-1 (IL-1); SPARC; thrombospondin; TSP; rho-A;
KW intracellular adhesion molecule-2; ICAM-2; von Willebrand Factor; VWF;
KW angiolipomas; inflammatory disorder; developmental disorder; wound;
KW menstrual disorder; cancer; rheumatoid arthritis; atherosclerosis;
KW diabetic retinopathy; restenosis; osteoporosis; cataract;
KW diabetes mellitus; glomerulonephritis; inflammatory glomerular disease;
KW vasculitis; retinopathy; liver fibrosis; haemoglobinopathy;
KW Crohn's disease.
XX
XX Homo sapiens.
OS
XX
XX WO200188122-A2.
PN
XX
XX 22-NOV-2001.
PD
XX
XX 15-MAY-2001; 2001WO-GB02140.
PE
XX
XX 16-MAY-2000; 2000US-204493P.
PR
XX
XX (GLAXO) GLAXO GROUP LTD.
PA
XX
XX McLaughlin F, Randi AM;
PI
XX
XX WPI: 2002-089855/12.
DR
XX
XX New modulator of a member of the Ets family of transcription factors,
PT Ery, which is not tumour necrosis factor- α or interleukin-1, for
PT treating cancer, rheumatoid arthritis, atherosclerosis, restenosis, and
PT osteoporosis
XX
XX
PS Disclosure; Page 53-54; 75pp; English.

Query Match	89.3%;	Score 2182;	DB 23;	Length 462;
Best Local Similarity	88.7%;	Pred. No. 5.4e-188;		
Matches 407;	Conservative 14;	Mismatches 24;	Indels 14;	Gaps 4

OY	2	ASTIEALSVS	EDOSLE	ECVAGSPH	LAKTEMT	ASSSEY	QOTSMSPRV	QOWLSOP	61
		1	1	1	1	1	1	1	
Db	9	AAHIEALSVS	EDOSLE	ECVAGSPH	LAKTEMT	ASSSDV	QOTSMSPRV	QOWLSOP	68
OY	62	ARVITKMEC	NPNOVGS	HNSPD	CSVAKG	CMVSSSDV	GMNYSYME	EKHIPPNN	TTTN
Db	69	ARVITKMEC	NPNOVGS	HNSPD	CSVAKG	CMVSSSDV	GMNYSYME	EKHIPPNN	TTTN
OY	122	ERRVYVPAD	PPLMTSDH	RQML	LEMAV	KRYGL	PVDYDILL	FOUIDOKEL	CKMTKDFORLTP
Db	129	ERRVYVPAD	PPLMTSDH	RQML	LEMAV	KRYGL	PVDYDILL	FOUIDOKEL	CKMTKDFORLTP
OY	182	SYNADILL	SHLHYLER	GATFI	EPFTSV	YPEATORIT	TRP-----	DLEPCO	ARRSAM
Db	189	SYNADILL	SHLHYLER	GATFI	EPFTSV	YPEATORIT	TRP-----	DLEPCO	ARRSAM
OY	234	TSHSHPT--	OSKATOP	SSSTV	PKTE	DORPOL	DYIOIG	FTSSRL	ANPGSGOIQLOMFLTEL
Db	244	TGHGHP	PTPOSKA	OPSP	ETVP	PKTE	DORPOL	DYIOIG	FTSSRLANPGSGOIQLOMFLTEL
OY	293	LSDSSNSC	ITMEGNGE	FKMTD	DEYVAR	MRG	GEKRSK	PNMNYD	ILSTRALRYVYKNTMTK
Db	304	LSDSSNSC	ITMEGNGE	FKMTD	DEYVAR	MRG	GEKRSK	PNMNYD	ILSTRALRYVYKNTMTK
OY	353	VHGKRYAK	FDFH	GIQAOL	QHPRESS	MYKYP	SDLPYMS	SYNAH	POKNAFEAPRPA
Db	364	VHGKRYAK	FDFH	GIQAOL	QHPRESS	MYKYP	SDLPYMS	SYNAH	POKNAFEAPRPA
OY	413	TSSSFEAP	PNPYMS	PTGCI	YPTN	RLRPA	NHPSH	LGITY	451
Db	424	TSSSFEAP	PNPYMS	PTGCI	YPTN	RLRPA	NHPSH	LGITY	462

RESULT 4

ID AAR44556 standard; protein; 452 AA

AC AAR44556;

DT 26-MAY-1994 (first entry)

Human HUM-FLI-1 gene product.

KW chromosomal translocation; chimeric; chimaeric; Ewing sarcoma;

KW primitive peripheral neuroectodermal tumour; human chromosome 11;

XX

XX
0000000000

2000

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841

2000

24 XX
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857

PI Plougastel B, Thomas G, Zucman J;
xx

DR WPI; 1993-386580/48.

XX
XX
Nov: 30 1967

sequence involved in chromosomal trans-location, also derived

of Ewing sarcoma and melanoma

ps Disclosure; fig /; 123pp; French
xy

[illegible]

OS	Homo sapiens.
XX	
PN	WO200157277-A2.
XX	
PD	09-AUG-2001.
XX	
PF	30-JAN-2001; 2001WO-US000669.
XX	
PR	04-FEB-2000; 2000US-0180312.
PR	26-MAY-2000; 2000US-0207456.
PR	30-JUN-2000; 2000US-0608408.
PR	03-AUG-2000; 2000US-0632366.
PR	21-SEP-2000; 2000US-0234687.
PR	27-SEP-2000; 2000US-0236359.
PR	04-OCT-2000; 2000GB-0024263.
XX	
PA	(MOLE-) MOLECULAR DYNAMICS INC.
XX	
PI	Penn SG, Hanzel DK, Chen W, Rank DR;
DR	WPI; 2001-483447/52.
XX	
PT	Human genome-derived single exon nucleic acid probes useful for analyzing gene expression in human fetal liver -
PS	Claim 27; SEQ ID NO 26847; 6399p + sequence listing; English.
XX	
CC	The invention relates to a single exon nucleic acid probe for measuring human gene expression in a sample derived from human foetal liver. The single exon nucleic acid probes may be used for predicting, measuring and displaying gene expression in samples derived from human fetal liver. The present sequence is a peptide encoded by a single exon nucleic acid probe of the invention.
CC	Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
CC	
XX	
SQ	Sequence 188 AA;
Query Match	38.8%; Score 948; DB 22; Length 188;
Best Local Similarity	95.5%; Pred. No. 3.2e-77;
Matches 169; Conservative	4; Mismatches 4; Indels 0; Gaps 0;
OY	275 LANPGSGOIQLMQFLLELLSDSSNSNCITWEGTNGEFKMTDPDEVARMRGKRKKPMNMY 334 : 12 LTLSSGGIOQLMQFLLELLSDSSNSCITWEGTNEFKMTDPDEVARMRGKRKKPMNMY 71
DB	335 DKLSPALRYYYDKNTMTVHGKRYAYKFDFHGIAGIQAOLPHPESSMKYRPPSDLRYMGSYH 394 : 72 DKLSPALRYYYDKNTMTVHGKRYAYKFDFHGIAGIQAOLPHPESSLKYRPPDLRYMGSYH 131
OY	395 AHPQKMNVAPRPPLRLPTYSSTFFAARPYPNNSPRGIGLYPTRLPAAMPHSLGYTY 451 : 132 AHPQKMNVAPRPPLRLPTYSSTFFAARPYPNNSPRGIGLYPTRLPTSHMPHSLGYTY 188
RESULT 8	
ABBI9649	
ID	ABBI9649 standard; Protein: 188 AA.
XX	
AC	ABBI9649;
XX	
DT	23-JAN-2002 (first entry)
DE	Protein #1648 encoded by probe for measuring heart cell gene expression.
XX	
KW	Human; gene expression; heart; microarray; vascular system; cardiovascular disease; hypertension; cardiac arrhythmia; congenital heart disease.
KW	
XX	
OS	Homo sapiens.
XX	
PN	WO200157277-A2.

```
XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US00666.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488899/53.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
XX hearts -
XX
XX Claim 15; SEQ ID No 21419; 530bp; English.
XX
XX The present invention relates to single exon nucleic acid probes for
XX measuring human gene expression in a sample derived from human heart (see
XX ABA21533-ABA41305). The present sequence is a protein encoded by one such
XX probe. The probes may be used for predicting, measuring and displaying
XX gene expression in samples derived from the human heart via microarrays.
XX By measuring gene expression, the probes are useful for predicting
XX diagnosing, grading, staging, monitoring and prognosing diseases of the
XX human heart and vascular system e.g. cardiovascular disease,
XX hypertension, cardiac arrhythmias and congenital heart disease.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 188 AA:
XX
XX Query Match 38.8%; Score 948; DB 22; Length 188;
XX Best Local Similarity 95.5%; Pred. No. 3.2e-77;
XX Matches 169; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
XX
XX QY 275 LANPGSGQIQLMQFLLELLSDSSNSNCITWEGNGEFGKMTDPDEVARRMGERKSKPNMY 334
XX | : |||||
XX DB 12 LFLSLGSGQIQLMQFLLELLSDSSNSNCITWEGNGEFGKMTDPDEVARRMGERKSKPNMY 71
XX
XX QY 335 DKLSRALRYYYDKNIMTKYHGKRYAKKFDPHGTAQALQPHPPSSLYKYPDDLPMGSH 394
XX |||||
XX DB 72 DKLSRALRYYYDKNIMTKYHGKRYAKKFDPHGTAQALQPHPPSSLYKYPDDLPMGSH 131
XX
XX QY 395 AHPQKMNFAVAPHPALPVTSSSFFAAPNPYWNSPGTGIVPNTLPAHMPSHLGTYY 451
XX |||||
XX DB 132 AHPQKMNFAVAPHPALPVTSSSFFAAPNPYWNSPGTGIVPNTLPAHMPSHLGTYY 188
XX
XX RESULT 9
XX AAM55005
XX ID AAM55005 standard; Protein; 188 AA.
XX
XX AC AAM55005;
XX
XX DT 05-NOV-2001 (first entry)
XX
XX DE Human brain expressed single exon probe encoded protein SEQ ID NO: 27110.
XX
XX KW Human; brain expressed exon; gene expression analysis; probe;
XX microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
XX epilepsy; cancer.
XX
XX OS Homo sapiens.
XX
XX PN WO200157275-A2.
```

```
XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US00667.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-483446/52.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
XX brains -
XX
XX Example 4; SEQ ID NO: 27110; 650bp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX brain. They can be used to measure gene expression in brain cell samples,
XX which may enable the diagnosis and improved treatment of nervous system
XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
XX epilepsy and cancers. The present sequence is a protein encoded by one of
XX the probes of the invention.
XX
XX Sequence 188 AA:
XX
XX Query Match 38.8%; Score 948; DB 22; Length 188;
XX Best Local Similarity 95.5%; Pred. No. 3.2e-77;
XX Matches 169; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
XX
XX QY 275 LANPGSGQIQLMQFLLELLSDSSNSNCITWEGNGEFGKMTDPDEVARRMGERKSKPNMY 334
XX | : |||||
XX DB 12 LFLSLGSGQIQLMQFLLELLSDSSNSNCITWEGNGEFGKMTDPDEVARRMGERKSKPNMY 71
XX
XX QY 335 DKLSRALRYYYDKNIMTKYHGKRYAKKFDPHGTAQALQPHPPSSLYKYPDDLPMGSH 394
XX |||||
XX DB 72 DKLSRALRYYYDKNIMTKYHGKRYAKKFDPHGTAQALQPHPPSSLYKYPDDLPMGSH 131
XX
XX QY 395 AHPQKMNFAVAPHPALPVTSSSFFAAPNPYWNSPGTGIVPNTLPAHMPSHLGTYY 451
XX |||||
XX DB 132 AHPQKMNFAVAPHPALPVTSSSFFAAPNPYWNSPGTGIVPNTLPAHMPSHLGTYY 188
XX
XX RESULT 10
XX AAM67391
XX ID AAM67391 standard; Protein; 188 AA.
XX
XX AC AAM67391;
XX
XX DT 06-NOV-2001 (first entry)
XX
XX DE Human bone marrow expressed probe encoded protein SEQ ID NO: 27697.
XX
XX KW Human; bone marrow expressed exon; gene expression analysis; probe;
XX microarray; cancer; leukaemia; lymphoma; myeloma.
XX
XX OS Homo sapiens.
XX
XX PN WO200157276-A2.
XX
XX DT 09-AUG-2001.
XX
XX PD 30-JAN-2001; 2001WO-US00668.
XX
XX PR 04-FEB-2000; 2000US-0180312.
```


Pt	Penn SG, Hanzel DK, Chen W, Rank DR;
Xx	
Dr	WPI: 2001-488697/53.
Xx	
Pt	Human genome-derived single exon nucleic acid probes useful for
Pt	analyzing gene expression in human placenta -
Xx	
Pt	Claim 27; SEQ ID No 27951; 654bp; English.
Xx	
Cc	The present invention relates to single exon nucleic acid probes (SENP;
Cc	see AAI315-AI37346). The present sequence is a peptide encoded by one
Cc	such probe. The probes are useful for producing a microarray for
Cc	predicting, measuring and displaying gene expression in samples derived
Cc	from human placenta. The probes are useful for antenatal diagnosis of
Cc	human genetic disorders.
Xx	
SQ	Sequence 188 AA:
	Query Match 38.8%; Score 948; DB 22; Length 188;
	Best Local Similarity 95.5%; Pred. No. 3.2e-77;
	Matches 169; Conservative 4; Mismatches 4; Indels 0; Gaps 0
Oy	275 LANPSSGQIQLMQFLLELLSDSSNSNCITWEGINGEFKMTDPDEVARRMGERKSKEPMNTY 334
Db	12 LPTSLSGGQIQLMQFLLELLSDSSNSCITWEGINGEFKMTDPDEVARRMGERKSKEPMNTY 71
Oy	335 DKLRSALRRYYKNKINTKYGKRKYAKFPDHGIQAOLOPHPRESSMYKYPSDLPRYMSYH 394
Db	72 DKLRSALRRYYKNKINTKYGKRKYAKFPDHGIQAOLOPHPRESSLYKKYPSDLPRMGSYH 131
Oy	395 AHPQKMNVAPRPPALPVTTSSSFPAAPNDPWNSPTGCIYPNTRLPAAHMPSHLGTYT 451
Db	132 AHPQKMNEVAPRPALPVTTSSSFPAAPNDPWNSPTGCIYPNTRLPSHMPSHLGTYT 188
	RESULT 13
ID	AAM02964
XX	AAM02964 standard; Protein; 188 AA.
XX	
XX	AAM02964;
XX	
Dt	09-OCT-2001 (first entry)
XX	
De	Peptide #1646 encoded by probe for measuring breast gene expression.
XX	
Kw	Probe; human; breast disease; breast cancer; development disorder;
Kw	inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX	
Os	Homo sapiens.
XX	
FN	WO200157270-A2.
XX	
PD	09-AUG-2001.
XX	
XX	29-JAN-2001; 2001WO-US00661.
XX	
Pr	04-FEB-2000; 2000US-0180312.
Pr	26-MAY-2000; 2000US-0207456.
Pr	30-JUN-2000; 2000US-0608408.
Pr	03-AUG-2000; 2000US-0632366.
Pr	21-SEP-2000; 2000US-0234687.
Pr	27-SEP-2000; 2000US-0236359.
Pr	04-OCT-2000; 2000GB-0024263.
XX	
Pa	(MOLE-) MOLECULAR DYNAMICS INC.
XX	
Pi	Penn SG, Hanzel DK, Chen W, Rank DR;
XX	
DR	WPI: 2001-476286/51.
XX	
Pt	Novel single exon nucleic acid probe used to measuring gene expression
Pt	in a human breast -
XX	

PS	Claim 27, SEQ ID No 11704; 322pp; English.
XX	
CC	The present invention relates to novel single exon nucleic acid probes
CC	(see A100010-A110067). The present sequence is a peptide encoded by one
CC	such probe. The probes are useful for measuring human gene expression in
CC	a human breast sample, where the probe hybridises at high stringency to a
CC	nucleic acid expressed in the human breast. The probes are useful for
CC	predicting, diagnosing, grading, staging, monitoring and prognosing
CC	diseases of the human breast, particularly those diseases with polygenic
CC	aetiology. The diseases include: breast cancer, disorders of development,
CC	inflammatory diseases of the breast, fibrocystic changes, proliferative
CC	breast disease and non-carcinoma tumours.
CC	Note: The sequence data for this patent did not form part of the printed
CC	specification, but was obtained in electronic format directly from WIPO
CC	at ftp.wipo.int/pub/published_pt_sequences .
XX	
SO	Sequence 188 AA:
Query Match	38.8%; Score 948; DB 22; Length 188;
Best Local Similarity	95.5%; Pred. No. 3.2e-77;
Matches 169; Conservative	4; Mismatches 4; Indels 0; Gaps 0
QY	275 LADGSGQIDLOMFLLELSDSSNSCTIWEETNGEFKMTDDEYARRWGERKKSPNNKY 334
Db	12 LTVSGSGQIDLOMFLLELSDSSNSCTIWEETNGEFKMTDDEYARRWGERKKSPNNKY 71
QY	335 DKLSRALRRYYDDNIMTKYHGKRYAKKPFPHGIAOALOPHPRESSMYKYPEDLRYMSYH 394
Db	72 DKLSRALRRYYDDNIMTKYHGKRYAKKPFPHGIAOALOPHPRESSLYKYPEDLRYMSYH 131
QY	395 AHPQKNFVAPHPALPVTSSEFFAAPNDPMYNSPTGGIYPMTRLPAAHMPSHLGTYY 451
Db	132 AHPQKNFVAPHPALPVTSSEFFAAPNDPMYNSPTGGIYPMTRLPESHMPSHLGTYY 188
RESULT 14	
ABG37017	
ID	ABG37017 standard; Peptide; 188 AA.
XX	
AC	ABG37017;
XX	
DT	19-AUG-2002 (first entry)
XX	
DE	
XX	
XX	Human peptide encoded by genome-derived single exon probe SEQ ID 26682.
KW	Human; single exon probe; asthma; lung cancer; COPD; ILD;
KW	chronic obstructive pulmonary disease; interstitial lung disease;
KW	familial idiopathic pulmonary fibrosis; neurofibromatosis;
KW	tubercular scleriosis; Gaucher's disease; Niemann-Pick disease;
KW	Hemansky-Rudik syndrome; sarcoidosis; pulmonary haemosiderosis;
KW	pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;
KW	pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
KW	primary ciliary dyskinesia; pulmonary hypertension;
KW	lysine membrane disease.
XX	
OS	Homo sapiens.
XX	
PN	WO200186003-A2.
XX	
PD	15-NOV-2001.
XX	
PF	30-JAN-2001; 2001WO-US00665.
PR	04-FEB-2000; 2000US-180312P.
PR	26-MAY-2000; 2000US-207456P.
PR	30-JUN-2000; 2000US-0608408.
PR	03-AUG-2000; 2000US-0632366.
PR	21-SEP-2000; 2000US-234687P.
PR	27-SEP-2000; 2000US-236599P.
PR	04-OCT-2000; 2000GB-0024263.
XX	
XX	(MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI: 2002-114183/15.
 XX
 PT Spatially-addressable set of single exon nucleic acid probes, used to
 PT measure gene expression in human lung samples -
 XX
 PS Claim 27; SEQ ID NO 26682; 634pp; English.
 XX
 CC The invention relates to a spatially-addressable set of single exon
 CC nucleic acid probes for measuring gene expression in a sample derived
 CC from human lung comprising single exon nucleic acid probes having one of
 CC 12614 nucleic acid sequences mentioned in the specification, or their
 CC complements or the 12367 open reading frames derived from the 12614
 CC probes. Also included are a microarray comprising the novel set of
 CC probes; the novel set of probes which hybridize at high stringency to a
 CC nucleic acid expressed in the human lung; measuring gene expression in a
 CC sample derived from human lung, comprising (a) contacting the array with
 CC a collection of detectably labeled nucleic acids derived from human lung
 CC mRNA, and (b) measuring the label detectably bound to each probe of
 CC the array; identifying exons in a eukaryotic genome, comprising
 CC (a) algorithmically predicting at least one exon from genomic sequences
 CC of the eukaryote; and (b) detecting specific hybridisation of detectably
 CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
 CC having a fragment identical to the predicted exon, the probe is included
 CC in the above mentioned microarray; assigning exons to a single gene,
 CC comprising (a) identifying exons from genomic sequence by the method
 CC above and (b) measuring the expression of each of the exons in several
 CC tissues and/or cell types using hybridisation to a single exon
 CC microarrays having a probe with the exon, where a common pattern of
 CC expression of the exons in the tissues and/or cell types indicates that
 CC the exons should be assigned to a single gene; a peptide comprising one
 CC of 12011 sequences, mentioned in the specification, or encoded by the
 CC probes/open reading frames (ORF). The probes are used for gene
 CC expression analysis, and for identifying exons in a gene, particularly
 CC using human lung derived mRNA and for the study of lung diseases
 CC such as asthma, lung cancer, chronic obstructive pulmonary disease
 CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary
 CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,
 CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary
 CC haemangioendothelioma, pulmonary histiocytosis, lymphangioleiomyomatosis,
 CC pulmonary alveolar proteinosis, Karsagen syndrome, fibrocystic
 CC and hyaline membrane disease. The present sequence is a peptide/protein
 CC encoded by a single exon probe of the invention.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX
 SQ Sequence 188 AA;
 Query Match 38.8%; Score 948; DB 23; Length 188;
 Best Local Similarity 95.5%; Pred. No. 3.2e-77;
 Matches 16; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
 QY 275 LANEQSGQIDLMQFLLELLSDSSNSNCITWEGTGEFKMTDPDEVARRWGRKSKPMNMY 334
 Db 12 LTLGSGQIDLMQFLLELLSDSSNSNCITWEGTGEFKMTDPDEVARRWGRKSKPMNMY 71
 QY 335 DKLSRALRYYYDKNIMKRYGKRAKYKFDHGIQALQPHRPRESSMKRYSDLPYMSYH 394
 Db 72 DKLSRALRYYYDKNIMKRYGKRAKYKFDHGIQALQPHRPRESSLKYKYSDDLPLYMSYH 131
 QY 395 AHPQKNFVAPHPALPVTSSFFAAPNPYNSPTGGIYPTRLPAAHMPDSLGTYY 451
 Db 132 AHPQKNFVAPHPALPVTSSFFAAPNPYNSPTGGIYPTRLPAAHMPDSLGTYY 188
 RESULT 15
 ABB31070
 ID ABB31070 standard; Peptide; 144 AA.

AC ABB31070;
 XX
 DT 01-FEB-2002 (first entry)
 XX
 DE Peptide #3721 encoded by breast cell single exon nucleic acid probe.
 XX
 KW Human; microarray; single exon probe; gene expression; breast;
 XX disease; cancer.
 XX
 OS Homo sapiens.
 XX
 PN WO200157271-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00662.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI: 2001-496933/54.
 XX
 XX New spatially-addressable set of single exon nucleic acid probes,
 PT useful for measuring gene expression in sample derived from human
 PT breast, comprises number of single exon nucleic acid probes -
 XX
 PS Claim 27; SEQ ID NO 14038; 327pp + sequence listing; English.
 XX
 CC The invention relates to a spatially-addressable set of single exon
 CC nucleic acid probes for measuring gene expression in a sample derived
 CC from human breast and BT 474 cells. The method involves contacting
 CC the probes with a collection of detectably labelled nucleic acids
 CC derived from mRNA of human breast, and then measuring the label
 CC bound to each probe of the microarray. The probes are useful for
 CC verifying the expression of regions of genomic DNA predicted to
 CC encode proteins. They are useful for gene discovery, and for
 CC determining predisposition and/or prognosing breast disease. Gene
 CC expression analysis is useful for assessing the toxicity of chemical
 CC agents on cells. The microarray of this invention presents a far greater
 CC diversity of probes for measuring gene expression, with far less bias
 CC than expressed sequence tag microarrays. The method is suitable for
 CC rapid production of functional information from genomic sequence. The
 CC present sequence is a peptide encoded by a single exon nucleic acid
 CC probe of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX
 SQ Sequence 144 AA;
 Query Match 32.5%; Score 794; DB 22; Length 144;
 Best Local Similarity 97.9%; Pred. No. 1.7e-53;
 Matches 14; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 292 LLSDSNSNCITWEGTGEFKMTDPDEVARRWGRKSKPMNMYDKLSRALRYYYDKNIMT 351
 Db 1 LLSDSNSNCITWEGTGEFKMTDPDEVARRWGRKSKPMNMYDKLSRALRYYYDKNIMT 60
 QY 352 KVHGKRAYAYKFDHGIQALQPHRPRESSMYKYSDDLPLYMSYHNAHPQKNFVAPHPALP 411
 Db 61 KVHGKRAYAYKFDHGIQALQPHRPRESSLKYKYSDDLPLYMSYHNAHPQKNFVAPHPALP 120
 QY 412 VTSSFFAAPNPYNSPTGGIYPTRLPAAHMPDSLGTYY 435
 ID VTSSFFAAPNPYNSPTGGIYPTRLPAAHMPDSLGTYY 435

Db 121 VTSSFFAABNPYWNPTGCIIPN 144

Search completed: November 9, 2002, 16:50:22
Job time : 55.4015 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 9, 2002, 10:20:16 ; Search time 1.19181 Seconds

(without alignments)
4631.768 Million cell updates/sec

Title: US-09-902-772-1_COPY_645_662

Perfect score: 18

Sequence: 1 ctcagagagagagagc 18

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_MA.*

1: /cgn2_6/ptodata1/ina/5A.COMB.seq.*

2: /cgn2_6/ptodata1/ina/5B.COMB.seq.*

3: /cgn2_6/ptodata1/ina/6A.COMB.seq.*

4: /cgn2_6/ptodata1/ina/6B.COMB.seq.*

5: /cgn2_6/ptodata1/ina/PCTUS.COMB.seq.*

6: /cgn2_6/ptodata1/ina/backfile1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18	100.0	1447	4	US-08-878-177-1
2	16.4	91.1	1956	3	US-08-313-553-6
3	16.4	91.1	1956	3	US-08-767-993-6
4	15.4	85.6	81001	4	US-09-750-580-1
5	15	83.3	4852	1	US-07-853-913-3
6	15	83.3	7874	4	US-09-780-175-96
7	14.8	82.2	2413	3	US-08-651-579-1
8	14.8	82.2	2921	3	US-08-904-032-2
9	14.8	82.2	11613	1	US-08-484-044-10
10	14.4	80.0	345	4	US-08-991-789A-154
11	14.4	80.0	345	4	US-09-062-451-154
12	14.4	80.0	345	4	US-09-598-126-154
13	14.4	80.0	2061	6	5171840-1
14	14.4	80.0	2061	6	5480796-1
15	14.4	80.0	3319	4	US-08-795-473B-2
16	14.4	80.0	3319	4	US-09-439-856-2
17	14.4	80.0	111282	4	US-09-754-250-3
18	14	77.8	2679	3	US-08-504-284-4
19	14	77.8	3348	4	US-09-302-620B-94
20	14	77.8	6433	4	US-08-927-219-128
21	14	77.8	11288	4	US-08-646-301A-1
22	14	77.8	11288	4	US-08-481-968A-4
23	14	77.8	11288	4	US-08-154-712B-4
24	13.8	76.7	1306	2	US-08-671-975A-1
25	13.8	76.7	1340	4	US-09-647-826-2
26	13.8	76.7	1345	4	US-09-372-339-1
27	13.8	76.7	1345	4	US-09-372-339-2

C 28	13.8	76.7	1345	4	US-09-144-367-3	Sequence 3, Appl
C 29	13.8	76.7	1366	1	US-08-091-519-12	Sequence 12, Appl
C 30	13.8	76.7	1366	1	US-08-442-043A-12	Sequence 12, Appl
C 31	13.8	76.7	1366	5	PCT-US91-03478-12	Sequence 12, Appl
C 32	13.8	76.7	1530	2	US-08-704-931-3	Sequence 3, Appl
C 33	13.8	76.7	1533	1	US-08-489-733-2	Sequence 2, Appl
C 34	13.8	76.7	1533	2	US-08-993-581B-2	Sequence 2, Appl
C 35	13.8	76.7	1533	4	US-09-134-078-8	Sequence 8, Appl
C 36	13.8	76.7	1610	5	US-08-889-108-7	Sequence 7, Appl
C 37	13.8	76.7	1610	5	PCT-US94-10358-7	Sequence 7, Appl
C 38	13.8	76.7	1656	4	US-09-522-217-106	Sequence 106, App
C 39	13.8	76.7	1687	1	US-08-656-984A-33	Sequence 33, Appl
C 40	13.8	76.7	2160	4	US-08-188-273A-1	Sequence 1, Appl
C 41	13.8	76.7	2162	4	US-09-351-198-1	Sequence 1, Appl
C 42	13.8	76.7	2162	4	US-09-113-426-1	Sequence 1, Appl
C 43	13.8	76.7	2399	2	US-08-969-106-9	Sequence 9, Appl
C 44	13.8	76.7	2421	1	US-08-063-552-1	Sequence 1, Appl
C 45	13.8	76.7	2421	5	PCT-US93-05704-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-878-177-1
Sequence 1, Application US/08878177
Patent No. 6294354
GENERAL INFORMATION:
APPLICANT: Chugai Seiyaku Kabushiki Kaisha, 5001, Iwamoto et
TITLE OF INVENTION: Cell Calcification Suppressing Proteins and Genes of
FILE REFERENCE: Chugai Seiyaku Kabushiki Kaisha 5001
CURRENT APPLICATION NUMBER: US/08/878.177
CURRENT FILING DATE: 1997-06-18
NUMBER OF SEQ. ID NOS: 7
SOFTWARE: PatentIn Ver. 2.0
SEQ. ID NO 1
LENGTH: 1447
TYPE: DNA
ORGANISM: C-11 gene, c-ery gene w/ deletion, chicken DNA
US-08-878-177-1

Query Match 100.0%; Score 18; DB 4; Length 1447;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCAGAGAGAGAGGCC 18
DB 645 CTCAGAGAGAGAGGCC 662

RESULT 2
US-08-313-553-6
Sequence 6, Application US/08313553
Patent No. 5641650
GENERAL INFORMATION:
APPLICANT: TURNER, George J.
APPLICANT: BETLACH, Mary C.
TITLE OF INVENTION: EXPRESSION OF HETEROLOGOUS POLYPEPTIDES
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Walter H. Dregger
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/313,553
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/038,662
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Dreger, Walter H.
REGISTRATION NUMBER: 24,190
REFERENCE/DOCKET NUMBER: A-57669/WHD
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1956 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 376..1812
FEATURE:
NAME/KEY: misc_feature
LOCATION: 376..414
OTHER INFORMATION: /note= "Bacteriorhodopsin
OTHER INFORMATION: pre-sequence."
FEATURE:
NAME/KEY: terminator
LOCATION: 1864..1866
OTHER INFORMATION: /note= "Bacteriorhodopsin stop
FEATURE:
NAME/KEY: mutation
LOCATION: replace(213, "")
OTHER INFORMATION: /note= "G to T mutation removes
OTHER INFORMATION: AlwNI restriction site."
FEATURE:
NAME/KEY: misc_feature
LOCATION: 427..435
OTHER INFORMATION: /note= "AlwNI cloning site."
FEATURE:
NAME/KEY: mutation
LOCATION: replace(930, "")
OTHER INFORMATION: /note= "G to A mutation removes
OTHER INFORMATION: AlwNI restriction site."
FEATURE:
NAME/KEY: mutation
LOCATION: replace(1179, "")
OTHER INFORMATION: /note= "T to A mutation removes
OTHER INFORMATION: AlwNI site."
FEATURE:
NAME/KEY: mutation
LOCATION: replace(1245, "")
OTHER INFORMATION: /note= "G to A mutation removes
OTHER INFORMATION: PstI restriction site."
FEATURE:
NAME/KEY: misc_signal
LOCATION: 374
OTHER INFORMATION: /note= "RNA start site."
FEATURE:
NAME/KEY: mutation
LOCATION: replace(1863, "")
OTHER INFORMATION: /note= "C to T mutation removes
OTHER INFORMATION: AlwNI restriction site."
FEATURE:
NAME/KEY: terminator
LOCATION: 1813..1815
OTHER INFORMATION: /note= "Muscarinic "OM1" stop
OTHER INFORMATION: codon."
US-08-313-553-6

```

```

Query Match          91.1%; Score 16.4; DB 1; Length 1956;
Best Local Similarity 94.4%; Pred. No. 27;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTCAGAGAGAGAGACC 18
DB 1305 CTCAGAGAGAGAGACC 1322

RESULT 3
US-08-767-993-6
; Sequence 6, Application US/08767993
; Patent No. 6010885
; GENERAL INFORMATION:
; APPLICANT: TURNER, George J.
; APPLICANT: BETLACH, Mary C.
; TITLE OF INVENTION: EXPRESSION OF HETEROLOGOUS POLYPEPTIDES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Walter H. Dreger
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/767,993
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Dreger, Walter H.
; REGISTRATION NUMBER: 24,190
; REFERENCE/DOCKET NUMBER: A-57669/WHD
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1956 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 376..1812
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 376..414
; OTHER INFORMATION: /note= "Bacteriorhodopsin
; OTHER INFORMATION: pre-sequence."
; FEATURE:
; NAME/KEY: terminator
; LOCATION: 1864..1866
; OTHER INFORMATION: /note= "Bacteriorhodopsin stop
; OTHER INFORMATION: codon."
; FEATURE:
; NAME/KEY: mutation
; LOCATION: replace(213, "")
; OTHER INFORMATION: /note= "G to T mutation removes
; OTHER INFORMATION: AlwNI restriction site."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 427..435
; OTHER INFORMATION: /note= "AlwNI cloning site."
; FEATURE:

```

```
NAME/KEY: mutation
LOCATION: replace(930, "")
OTHER INFORMATION: /note="G to A mutation removes
OTHER INFORMATION: AlwNI restriction site."
FEATURE:
NAME/KEY: mutation
LOCATION: replace(1179, "")
OTHER INFORMATION: /note="T to A mutation removes
OTHER INFORMATION: AlwNI site."
FEATURE:
NAME/KEY: mutation
LOCATION: replace(1245, "")
OTHER INFORMATION: /note="G to A mutation removes
OTHER INFORMATION: PstI restriction site."
FEATURE:
NAME/KEY: misc_signal
LOCATION: 374
OTHER INFORMATION: /note="RNA start site."
FEATURE:
NAME/KEY: mutation
LOCATION: replace(1863, "")
OTHER INFORMATION: /note="C to T mutation removes
OTHER INFORMATION: AlwNI restriction site."
FEATURE:
NAME/KEY: terminator
LOCATION: 1813..1815
OTHER INFORMATION: /note="Muscarinic 'OM1' stop
OTHER INFORMATION: codon."
US-08-767-993-6

Query Match      91.1%; Score 16.4; DB 3; Length 1956;
Best Local Similarity 94.4%; Pred. No. 27;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy      1 CTCAGAGAGAGAGGCC 18
Db      1305 CTCAGAGAGAGAGGCC 1322

RESULT 4
US-09-750-580-1
; Sequence 1, Application US/09750580
; Patent No. 6455280
; GENERAL INFORMATION:
; APPLICANT: Yen, Frances
; APPLICANT: Denison, Blake
; APPLICANT: Bour, Barbara
; APPLICANT: Bihain, Bernard
; APPLICANT: Dumas Milne Edwards, Jean-Baptiste
; APPLICANT: Duclet, Aymeric
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Edbets-Reed, Dana
; APPLICANT: Salter-Cid, Luisa
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITING NEOPLASTIC CELL GROWTH
; FILE REFERENCE: 89.052.CIP
; CURRENT APPLICATION NUMBER: US/09/750.580
; CURRENT FILING DATE: 2000-12-28
; PRIOR APPLICATION NUMBER: US 09/599.362
; PRIOR FILING DATE: 2000-06-21
; PRIOR APPLICATION NUMBER: PCT/IB00/0101
; PRIOR FILING DATE: 2000-06-21
; PRIOR APPLICATION NUMBER: PCT/IB99/02058
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: US 49/469/099
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: US 60/113.686
; PRIOR FILING DATE: 1998-12-22
; PRIOR APPLICATION NUMBER: US 60/141.032
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patent.pm
; SEQ ID NO 1
; LENGTH: 81001
```

```
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10946..12946
OTHER INFORMATION: 5' regulatory region
NAME/KEY: exon
LOCATION: 12947..12958
OTHER INFORMATION: exon 1
NAME/KEY: exon
LOCATION: 13470..13526
OTHER INFORMATION: exon 2
NAME/KEY: exon
LOCATION: 13641..13752
OTHER INFORMATION: exon 3
NAME/KEY: exon
LOCATION: 14271..15368
OTHER INFORMATION: exon 4
NAME/KEY: misc_feature
LOCATION: 15969..17969
OTHER INFORMATION: 3' regulatory region
NAME/KEY: allele
LOCATION: 1239
OTHER INFORMATION: 20-828-311 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 12347
OTHER INFORMATION: 17-42-319 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 15241
OTHER INFORMATION: 17-41-250 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 42218
OTHER INFORMATION: 20-841-149 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 45442
OTHER INFORMATION: 20-842-115 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 77058
OTHER INFORMATION: 20-853-415 : polymorphic base C or T
NAME/KEY: primer_bind
LOCATION: 929..949
OTHER INFORMATION: 20-828-.pu
NAME/KEY: primer_bind
LOCATION: 1357..1377
OTHER INFORMATION: 20-828-.rp complement
NAME/KEY: primer_bind
LOCATION: 12029..12050
OTHER INFORMATION: 17-42-.pu
NAME/KEY: primer_bind
LOCATION: 12581..12603
OTHER INFORMATION: 17-42-.rp complement
NAME/KEY: primer_bind
LOCATION: 14992..15012
OTHER INFORMATION: 17-41-.pu
NAME/KEY: primer_bind
LOCATION: 15460..15482
OTHER INFORMATION: 17-41-.rp complement
NAME/KEY: primer_bind
LOCATION: 42070..42090
OTHER INFORMATION: 20-841-.pu
NAME/KEY: primer_bind
LOCATION: 42572..42591
OTHER INFORMATION: 20-841-.rp complement
NAME/KEY: primer_bind
LOCATION: 45328..45347
OTHER INFORMATION: 20-842-.pu
NAME/KEY: primer_bind
LOCATION: 45863..45883
OTHER INFORMATION: 20-842-.rp complement
NAME/KEY: primer_bind
LOCATION: 76644..76664
OTHER INFORMATION: 20-853-.pu
NAME/KEY: primer_bind
```

LOCATION: 77166..77185
OTHER INFORMATION: 20-853.rp complement
NAME/KEY: primer_bind
LOCATION: 1220..1238
OTHER INFORMATION: 20-828-311.mis
NAME/KEY: primer_bind
LOCATION: 1240..1258
OTHER INFORMATION: 20-828-311.mis complement
NAME/KEY: primer_bind
LOCATION: 12328..12346
OTHER INFORMATION: 17-42-319.mis
NAME/KEY: primer_bind
LOCATION: 12348..12366
OTHER INFORMATION: 17-42-319.mis complement
NAME/KEY: primer_bind
LOCATION: 15222..15240
OTHER INFORMATION: 17-41-250.mis
NAME/KEY: primer_bind
LOCATION: 15242..15260
OTHER INFORMATION: 17-41-250.mis complement
NAME/KEY: primer_bind
LOCATION: 42199..42217
OTHER INFORMATION: 20-841-149.mis
NAME/KEY: primer_bind
LOCATION: 42219..42237
OTHER INFORMATION: 20-841-149.mis complement
NAME/KEY: primer_bind
LOCATION: 45423..45441
OTHER INFORMATION: 20-842-115.mis
NAME/KEY: primer_bind
LOCATION: 45443..45461
OTHER INFORMATION: 20-842-115.mis complement
NAME/KEY: primer_bind
LOCATION: 77039..77057
OTHER INFORMATION: 20-853-415.mis
NAME/KEY: primer_bind
LOCATION: 77059..77077
OTHER INFORMATION: 20-853-415.mis complement
NAME/KEY: misc_bind
LOCATION: 1227..1251
OTHER INFORMATION: 20-828-311.probe
NAME/KEY: misc_bind
LOCATION: 12335..12359
OTHER INFORMATION: 17-42-319.probe
NAME/KEY: misc_bind
LOCATION: 15229..15253
OTHER INFORMATION: 17-41-250.probe
NAME/KEY: misc_bind
LOCATION: 42206..42230
OTHER INFORMATION: 20-841-149.probe
NAME/KEY: misc_bind
LOCATION: 45430..45454
OTHER INFORMATION: 20-842-115.probe
NAME/KEY: misc_bind
LOCATION: 77046..77070
OTHER INFORMATION: 20-853-415.probe
US-09-750-380-1

Query Match 85.6%; Score 15.4; DB 4; Length 81001;
Best Local Similarity 94.1%; Pred. No. 92;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 CTCAGAGAGAGAGC 17
DB 47018 CTCAGAGAGAGATGAGC 47034
RESULT 5
US-07-853-913-3/c
Sequence 3, Application US/07853913
Patent No. 5338839
GENERAL INFORMATION:
APPLICANT: McKay, Ronald D.G.

APPLICANT: Lendahl, Urban
TITLE OF INVENTION: Nestin Expression As An Indicator of
TITLE OF INVENTION: Neuroepithelial Tumors
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Millitia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/853,913
FILING DATE: 19920319
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/660,412
FILING DATE: 22-FEB-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/603,803
FILING DATE: 25-OCT-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/201,762
FILING DATE: 02-JUN-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/180,548
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: MIT-4641AAAA
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-6240
TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 4852 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-07-853-913-3

Query Match 83.3%; Score 15; DB 1; Length 4852;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 CAGAGAGAGAGC 17
DB 1239 CAGAGAGAGAGC 1225

RESULT 6
US-09-780-175-96/c
Sequence 96, Application US/09780175
Patent No. 6440738
GENERAL INFORMATION:
APPLICANT: Robert McKay
APPLICANT: Susan M. Freier
APPLICANT: Jacqueline Wyatt
TITLE OF INVENTION: ANTISENSE MODULATION OF CASEIN KINASE 2-BETA EXPRESSION
FILE REFERENCE: KRS-0164
CURRENT APPLICATION NUMBER: US/09/780,175
CURRENT FILING DATE: 2001-02-08
NUMBER OF SEQ ID NOS: 154
SEQ ID NO 96
LENGTH: 7874
TYPE: DNA
ORGANISM: Mus musculus

FEATURE:
; NAME/KEY: CDS
; LOCATION: (2169)...(2297)
; NAME/KEY: CDS
; LOCATION: (4112)...(4150)
; NAME/KEY: CDS
; LOCATION: (2833)...(2913)
; NAME/KEY: CDS
; LOCATION: (4576)...(4614)
; NAME/KEY: CDS
; LOCATION: (4827)...(4994)
; NAME/KEY: CDS
; LOCATION: (6267)...(6431)
US-09-780-175-96

Query Match 83.3%; Score 15; DB 4; Length 7874;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CTCAGAGAGAGAGA 15
Db 5597 CTCAGAGAGAGAGA 5583

RESULT 7
US-08-651-579-1/c
; Sequence 1, Application US/08651579
; Patent No. 5912141
; GENERAL INFORMATION:
; APPLICANT: BROJATSCH, JURGEN
; APPLICANT: NAUGHTON, JOHN
; APPLICANT: YOUNG, JOHN A. T.
; TITLE OF INVENTION: TUMOR VIRUS SUSCEPTIBILITY GENES AND GENE
; TITLE OF INVENTION: PRODUCTS, AND USES RELATED THERETO
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHYE & COCKFIELD
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII (text)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/651,579
; FILING DATE: 22 MAY 1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Vincent, Matthew P.
; REGISTRATION NUMBER: 36,709
; REFERENCE/DOCKET NUMBER: HMT-019
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2413 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 215..1318
US-08-651-579-1

Query Match 82.2%; Score 14.8; DB 2; Length 2413;
Best Local Similarity 88.9%; Pred. No. 1.5e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 CTCAGAGAGAGAGACC 18
Db 1432 CTCAGAGAGAGAGAGACC 1415

RESULT 8
US-08-904-032-2/c
; Sequence 2, Application US/08904032
; Patent No. 6004778
; GENERAL INFORMATION:
; APPLICANT: Bauman, Olga
; APPLICANT: Lal, Preeti
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: EMBRYOGENESIS PROTEIN
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/904,032
; FILING DATE: Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0359 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2921 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: BLADNOT03
; CLONE: 1602555
US-08-904-032-2

Query Match 82.2%; Score 14.8; DB 3; Length 2921;
Best Local Similarity 88.9%; Pred. No. 1.5e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 CTCAGAGAGAGAGACC 18
Db 2040 CTCAGAGAGAGAGACC 2023

RESULT 9
US-08-484-044-10/c
; Sequence 10, Application US/08484044
; Patent No. 5552282
; GENERAL INFORMATION:
; APPLICANT: Caskey, C. T.
; APPLICANT: Fu, Ying-Hui
; APPLICANT: Friedman, David L.
; APPLICANT: Pizzuti, Antonio
; APPLICANT: Fenwick, Raymond G.

TITLE OF INVENTION: Diagnosis of Myotonic Muscular Dystrophy
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fulbright & Jaworski, L.L.P.
STREET: 1301 McKinney, Suite 5100
CITY: Houston
STATE: Texas
COUNTRY: U.S.A.
ZIP: 77010-3095
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,044
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/019,940
FILING DATE: 19-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: Paul, Thomas D.
REGISTRATION NUMBER: 32,714
REFERENCE/DOCKET NUMBER: D-5443
TELECOMMUNICATION INFORMATION:
TELEPHONE: 713/651-5325
TELEFAX: 713/651-5246
TELEX: 762829
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 11613 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-484-044-10

Query Match 82.2%; Score 14.8; DB 1; Length 11613;
Best Local Similarity 88.9%; Pred. No. 1.6e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTCAGAGAGAGAGAGCC 18
||||| |||||||
DB 4318 CTCAGAGAGAGAGAGCC 4301

RESULT 10
US-08-991-789A-154
Sequence 154, Application US/08991789A
Patent No. 6225054
GENERAL INFORMATION:
APPLICANT: Fridakis, Tony N.
Smith, John M.
Reed, Steven G.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
TREATMENT AND DIAGNOSIS OF BREAST CANCER
NUMBER OF SEQUENCES: 292
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed IP Law Group
STREET: 701 Fifth Avenue, Suite 6300
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/991,789A
FILING DATE: 11-Dec-1997

CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Potler, Jane E. R.
REGISTRATION NUMBER: 33,332
REFERENCE/DOCKET NUMBER: 210121.419C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 154:
SEQUENCE CHARACTERISTICS:
LENGTH: 345 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-991-789A-154

Query Match 80.0%; Score 14.4; DB 4; Length 345;
Best Local Similarity 93.8%; Pred. No. 2.1e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTCAGAGAGAGAGAG 16
||||| |||||||
DB 133 CTCAGAGAGAGAGAG 148

RESULT 11
US-09-062-451-154
Sequence 154, Application US/09062451
Patent No. 6344550
GENERAL INFORMATION:
APPLICANT: Fridakis, Tony N.
Smith, John M.
Reed, Steven G.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
TREATMENT AND DIAGNOSIS OF BREAST CANCER
NUMBER OF SEQUENCES: 297
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/062,451
FILING DATE: 04-APR-1997
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.419C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 154:
SEQUENCE CHARACTERISTICS:
LENGTH: 345 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-062-451-154

Query Match 80.0%; Score 14.4; DB 4; Length 345;
Best Local Similarity 93.8%; Pred. No. 2.1e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTCAGAGAGAGAGAG 16

Db 133 CTCAGAGAGGAGGAG 148

RESULT 12

US-09-598-326-154

Sequence 154, Application US/09598326

Patent No. 6423496

GENERAL INFORMATION:

APPLICANT: Fridakis, Tony N.

Smith, John M.

Reed, Steven G.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
TREATMENT AND DIAGNOSIS OF BREAST CANCER

NUMBER OF SEQUENCES: 247

CORRESPONDENCE ADDRESS:

ADDRESSEE: Seed Intellectual Property Law Group PLLC

STREET: 701 Fifth Avenue, Suite 6300

CITY: Seattle

STATE: Washington

COUNTRY: USA

ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/598.326

FILING DATE: 20-Jun-2000

CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Potter, Jane E.R.

REGISTRATION NUMBER: 33,332

REFERENCE/DOCKET NUMBER: 210121.419D1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 154:

SEQUENCE CHARACTERISTICS:

LENGTH: 345 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 154:

US-09-598-326-154

Query Match 80.0%; Score 14.4; DB 4; Length 345;

Best Local Similarity 93.8%; Pred. No. 2.1e+02;

Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTCAGAGAGGAGGAG 16
Db 133 CTCAGAGAGGAGGAG 148

RESULT 13

5171840-1

Patent No. 5171840

APPLICANT: KISHIMOTO, TADAMITSU

TITLE OF INVENTION: RECEPTOR PROTEIN FOR HUMAN B CELL

STIMULATORY FACTOR-2

NUMBER OF SEQUENCES: 11

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/298.694

FILING DATE: 19-JAN-1989

SEQ ID NO: 1:

LENGTH: 2061

5171840-1

Query Match 80.0%; Score 14.4; DB 6; Length 2061;

Best Local Similarity 93.8%; Pred. No. 2.2e+02;

Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 CAGAGAGAGAGGCC 18
Db 41 CAGAGGAGAGAGGCC 56

RESULT 14

5480796-1

Patent No. 5480796

APPLICANT: KISHIMOTO, TADAMITSU

TITLE OF INVENTION: ANTIBODIES AGAINST THE RECEPTOR PROTEIN

FOR HUMAN B CELL STIMULATORY FACTOR-2

NUMBER OF SEQUENCES: 8

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/907.650

FILING DATE: 02-JUL-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 298,694

FILING DATE: 19-JAN-1989

SEQ ID NO: 1:

LENGTH: 2061

5480796-1

Query Match 80.0%; Score 14.4; DB 6; Length 2061;

Best Local Similarity 93.8%; Pred. No. 2.2e+02;

Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 CAGAGAGAGAGGCC 18
Db 41 CAGAGGAGAGAGGCC 56

RESULT 15

US-08-795-473B-2

Sequence 2, Application US/08795473B

Patent No. 6217858

GENERAL INFORMATION:

APPLICANT: Galun, Elhan

APPLICANT: Nahot, Orit

APPLICANT: Blum, Herbert E.

TITLE OF INVENTION: A Pharmaceutical Composition for Treating

HEPATITIS B VIRUS (HBV) INFECTION

NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:

ADDRESSEE: Davidson, Davidson and Kappel, LLC

STREET: 1140 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 inch disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: MS-DOS EDITOR

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/795.473B

FILING DATE: 11-FEB-1997

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Davidson, Clifford M.

REGISTRATION NUMBER: 32,728

REFERENCE/DOCKET NUMBER: 963.1007

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212)-997-1028

TELEFAX: (212)-997-1037

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 3319 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: unknown

US-08-795-473B-2

Query Match 80.0%; Score 14.4; DB 4; Length 3319;
Best Local Similarity 93.8%; Pred. No. 2.3e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 3 CAGAGAGAGAGAGCC 18
Db 232 CAGAGGAGAGAGAGCC 247

Search completed: November 9, 2002, 12:30:07
Job time : 20.1918 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 9, 2002, 11:41:36 ; Search time 1.40068 Seconds
(without alignments)
4559.304 Million cell updates/sec

Title: US-09-902-772-1_COPY-645_662
Perfect score: 18
Sequence: 1 ctcaagagagagagagac 18

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 320260 seqs, 177392727 residues

Total number of hits satisfying chosen parameters: 640520

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
Published_Applications_NA: *
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq: *
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq: *
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq: *
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq: *
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq: *
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq: *
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq: *
8: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq: *
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq: *
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq: *
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq: *
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq: *
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq: *
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	15.4	85.6	170	10	US-09-864-761-25721
2	15.4	85.6	401	10	US-09-960-352-966
3	15.4	85.6	506	10	US-09-917-800A-944
4	15.4	85.6	583	10	US-09-864-761-9095
5	15.4	85.6	81001	10	US-09-751-877-1
6	15	83.3	171	10	US-09-878-574-5808
7	15	83.3	258	10	US-09-563-817-31
8	15	83.3	267	10	US-09-733-607-7
9	15	83.3	272	10	US-09-878-574-7091
10	15	83.3	283	10	US-09-878-574-11646
11	15	83.3	481	10	US-09-864-761-11469
12	15	83.3	547	10	US-09-824-401-19
13	15	83.3	5340	10	US-09-960-253-74
14	14.8	82.2	183	10	US-09-867-701-8586
15	14.8	82.2	236	10	US-09-960-352-12042
16	14.8	82.2	297	10	US-09-867-701-1009
17	14.8	82.2	507	10	US-09-867-701-988
18	14.8	82.2	597	10	US-09-833-381-360
19	14.8	82.2	1530	10	US-09-989-002-1

C 20	14.8	82.2	2094	10	US-09-925-297-246	Sequence 246, App
C 21	14.8	82.2	2229	10	US-09-822-830A-143	Sequence 143, App
C 22	14.8	82.2	2615	12	US-10-052-586-523	Sequence 523, App
C 23	14.8	82.2	2910	10	US-09-989-002-10	Sequence 10, App1
C 24	14.8	82.2	13747	10	US-09-964-824A-287	Sequence 287, App
C 25	14.8	82.2	167343	10	US-09-962-436-281	Sequence 281, App
C 26	14.8	82.2	167343	10	US-09-964-824A-273	Sequence 273, App
C 27	14.8	82.2	174424	10	US-09-967-768A-314	Sequence 314, App
C 28	14.8	82.2	302250	10	US-09-962-832-154	Sequence 154, App
C 29	14.4	80.0	241	10	US-09-604-878A-346	Sequence 346, App
C 30	14.4	80.0	241	12	US-10-007-805-346	Sequence 346, App
C 31	14.4	80.0	269	10	US-09-878-574-1267	Sequence 1267, A
C 32	14.4	80.0	271	10	US-09-878-574-8668	Sequence 8668, Ap
C 33	14.4	80.0	345	10	US-09-810-936-154	Sequence 154, App
C 34	14.4	80.0	345	10	US-09-429-755-154	Sequence 154, App
C 35	14.4	80.0	357	10	US-09-834-975-422	Sequence 422, App
C 36	14.4	80.0	401	10	US-09-795-668-733	Sequence 733, App
C 37	14.4	80.0	401	10	US-09-795-686-733	Sequence 733, App
C 38	14.4	80.0	429	10	US-09-878-574-3246	Sequence 3246, Ap
C 39	14.4	80.0	461	10	US-09-728-445-14	Sequence 14, App1
C 40	14.4	80.0	479	10	US-09-864-761-501	Sequence 501, App
C 41	14.4	80.0	479	10	US-09-864-761-17297	Sequence 17297, A
C 42	14.4	80.0	493	9	US-09-978-295A-291	Sequence 291, App
C 43	14.4	80.0	559	10	US-09-864-761-9133	Sequence 9133, Ap
C 44	14.4	80.0	603	10	US-09-974-300-6166	Sequence 6166, Ap
C 45	14.4	80.0	888	9	US-09-938-842A-2493	Sequence 2493, Ap

ALIGNMENTS

RESULT 1
US-09-864-761-25721
Sequence 25721, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
FILE REFERENCE: Aeomica-X-1
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US/09/864,761
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30

```

1  PRIOR APPLICATION NUMBER: PCT/US01/00670
2  PRIOR FILING DATE: 2001-01-30
3  PRIOR APPLICATION NUMBER: US 60/234, 687
4  PRIOR FILING DATE: 2000-09-21
5  PRIOR APPLICATION NUMBER: US 09/608, 408
6  PRIOR FILING DATE: 2000-06-30
7  PRIOR APPLICATION NUMBER: US 09/774, 203
8  PRIOR FILING DATE: 2001-01-29
9  NUMBER OF SEQ ID NOS: 49117
10 SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
11 SEQ ID NO 25721
12 LENGTH: 170
13 TYPE: DNA
14 ORGANISM: Homo sapiens
15 FEATURE:
16 OTHER INFORMATION: MAP TO AC007539.8
17 OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 6.5
18 OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 5.9
19 OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 7.9
20 OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 9.7
21 OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.9
22 OTHER INFORMATION: EXPRESSED IN BELA, SIGNAL = 4.6
23 OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 7.4
24 OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.4
25 OTHER INFORMATION: NT HIT: AFI56495.1, EVALUE 9.90e-02
26 OTHER INFORMATION: EST_HUMAN HIT: BT24974.1, EVALUE 3.70e-02
27 OTHER INFORMATION: SWISSPROT HIT: P40024, EVALUE 7.80e-01
28 US-09-864 -761-25721

```

Query Match	85.6%;	Score 15.4;	DB 10;	Length 170;
Best Local Similarity	94.1%;	Pred. No. 1.1e+02;		
Matches 16; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0;

QY	1	CTCAGAGAGAGAGGAGC	17
Db	144	CACAGAGAGAGAGGAGC	160

```

RESULT 2
US-09-960-352-966
: Sequence 966: Application US/09960352
: Patent No. US20020137139A1
: GENERAL INFORMATION:
: APPLICANT: Warren, Wesley C.
: APPLICANT: Tao, Nengping
: APPLICANT: Byatt, John C.
: APPLICANT: Mathialagan, Nagappan
: TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
: TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
: FILE REFERENCE: 16511.006/37-21(10298)C
: CURRENT APPLICATION NUMBER: US/09/960.352
: NUMBER OF SEQ ID NOS: 15112
: SEQ ID NO 966
: LENGTH: 401
: TYPE: DNA
: ORGANISM: Bos taurus
: OTHER INFORMATION: Clone ID: 05-LIB188-020-01-E1-B1
US-09-960-352-966

```

Query Match	85.68;	Score 15.4;	DB 10;	Length 401;
Best Local Similarity	94.18;	Pred. No. 1.2e+02;		
Matches 16; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0

QY	1	CTCAGAGAGAGAGAGC	17
Db	350	CTCAGGAGAGAGAGC	366

RESULT 3
US-09-917-800A-944
; Sequence 944, Application US/09917800A
; Patent No. US20020119462A1

GENERAL INFORMATION:
 APPLICANT: Mendrick, Donna
 APPLICANT: Porter, Mark
 APPLICANT: Johnson, Kory
 APPLICANT: Castle, Arthur
 APPLICANT: Elashoff, Michael
 APPLICANT: Gene Logic, Inc.
 TITLE OF INVENTION: Molecular Toxicology Modeling
 FILE REFERENCE: 44921-5038-US
 CURRENT APPLICATION NUMBER: US/09/917,800A
 CURRENT FILING DATE: 2001-07-31
 PRIOR APPLICATION NUMBER: US 60/222,040
 PRIOR FILING DATE: 2000-07-31
 PRIOR APPLICATION NUMBER: US 60/222,880
 PRIOR FILING DATE: 2000-11-02
 PRIOR APPLICATION NUMBER: US 60/290,029
 PRIOR FILING DATE: 2001-05-11
 PRIOR APPLICATION NUMBER: US 60/290,645
 PRIOR FILING DATE: 2001-05-15
 PRIOR APPLICATION NUMBER: US 60/292,336
 PRIOR FILING DATE: 2001-05-22
 PRIOR APPLICATION NUMBER: US 60/295,798
 PRIOR FILING DATE: 2001-06-06
 PRIOR APPLICATION NUMBER: US 60/297,457
 PRIOR FILING DATE: 2001-06-13
 PRIOR APPLICATION NUMBER: US 60/298,884
 PRIOR FILING DATE: 2001-06-19
 PRIOR APPLICATION NUMBER: US 60/303,459
 PRIOR FILING DATE: 2001-07-09
 NUMBER OF SEQ ID NOS: 1740
 SOFTWARE: Patentln Ver. 2.1
 SEQ ID NO 944
 LENGTH: 506
 TYPE: DNA
 ORGANISM: Rattus norvegicus
 FEATURE:
 OTHER INFORMATION: Genbank Accession No. US20020119462n1 AII175790
 US-09-917-800A-944

Query Match	85.6%	Score 15.4;	DB 10;	Length 506;
Best Local Similarity	94.1%	Pred. No. 1.2e+02;		
Matches 16; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0;

QY	1	CTCAGAGAGAGAGGC	17
Db	369	CTCAGAGAGAGAGGC	385

```

RESULT 4
US-09-864-761-9095
Sequence 9095, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
FILE REFERENCE: Aecomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666

```

PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 9095
LENGTH: 583
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC007539.8
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 6.5
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 5.9
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 7.9
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 9.7
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.9
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4.6
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 7.4
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.4
US-09-864-761-9095
Query Match 85.6%; Score 15.4; DB 10; Length 583;
Best Local Similarity 94.1%; Pred. No. 1.2e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
CY 1 CTCAGAGAGAGAGAGC 17
DB 437 CACAGAGAGAGAGAGC 453
RESULT 5
US-09-751-877-1
Sequence 1, Application US/09751877
Patent No. US20020142949a1
GENERAL INFORMATION:
APPLICANT: Yen, Frances
APPLICANT: Denison, Blake
APPLICANT: Bour, Barbara
APPLICANT: Bihain, Bernard
APPLICANT: Dumast Milne Edwards, Jean-Baptiste
APPLICANT: Duclert, Aymeric
APPLICANT: Bougueleret, Lydie
APPLICANT: Ebbels-Reed, Dana
APPLICANT: Salter-Cid, Luisa
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITTING NEOPLASTIC CELL GROWTH
FILE REFERENCE: 89, US3, REG
CURRENT APPLICATION NUMBER: US/09/751,877
CURRENT FILING DATE: 2000-12-28
NUMBER OF SEQ ID NOS: 6
SOFTWARE: Patent.pm
SEQ ID NO 1

LENGTH: 81001
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
LOCATION: 10946..12946
OTHER INFORMATION: 5'regulatory region
NAME/KEY: exon
LOCATION: 12947..12958
OTHER INFORMATION: exon 1
NAME/KEY: exon
LOCATION: 13470..13526
OTHER INFORMATION: exon 2
NAME/KEY: exon
LOCATION: 13641..13752
OTHER INFORMATION: exon 3
NAME/KEY: exon
LOCATION: 14271..15968
OTHER INFORMATION: exon 4
NAME/KEY: misc.feature
LOCATION: 15969..17969
OTHER INFORMATION: 3'regulatory region
NAME/KEY: allele
LOCATION: 1239
OTHER INFORMATION: 20-828-311 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 12347
OTHER INFORMATION: 17-42-319 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 15241
OTHER INFORMATION: 17-41-250 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 42218
OTHER INFORMATION: 20-841-149 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 45442
OTHER INFORMATION: 20-842-115 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 77058
OTHER INFORMATION: 20-853-415 : polymorphic base C or T
NAME/KEY: primer_bind
LOCATION: 929..949
OTHER INFORMATION: 20-828..pu
NAME/KEY: primer_bind
LOCATION: 1357..1377
OTHER INFORMATION: 20-828..rp complement
NAME/KEY: primer_bind
LOCATION: 12029..12050
OTHER INFORMATION: 17-42..pu
NAME/KEY: primer_bind
LOCATION: 12581..12603
OTHER INFORMATION: 17-42..rp complement
NAME/KEY: primer_bind
LOCATION: 14992..15012
OTHER INFORMATION: 17-41..pu
NAME/KEY: primer_bind
LOCATION: 15460..15482
OTHER INFORMATION: 17-41..rp complement
NAME/KEY: primer_bind
LOCATION: 42070..42090
OTHER INFORMATION: 20-841..pu
NAME/KEY: primer_bind
LOCATION: 42572..42591
OTHER INFORMATION: 20-841..rp complement
NAME/KEY: primer_bind
LOCATION: 45328..45347
OTHER INFORMATION: 20-842..pu
NAME/KEY: primer_bind
LOCATION: 45863..45883
OTHER INFORMATION: 20-842..rp complement
NAME/KEY: primer_bind
LOCATION: 76644..76664
OTHER INFORMATION: 20-853..pu

```
NAME/KEY: primer_bind
LOCATION: 77166..77185
OTHER INFORMATION: 20-853-rp complement
NAME/KEY: primer_bind
LOCATION: 1220..1238
OTHER INFORMATION: 20-828-311.mis
NAME/KEY: primer_bind
LOCATION: 1240..1258
OTHER INFORMATION: 20-828-311.mis complement
NAME/KEY: primer_bind
LOCATION: 12328..12346
OTHER INFORMATION: 17-42-319.mis
NAME/KEY: primer_bind
LOCATION: 12348..12366
OTHER INFORMATION: 17-42-319.mis complement
NAME/KEY: primer_bind
LOCATION: 15222..15240
OTHER INFORMATION: 17-41-250.mis
NAME/KEY: primer_bind
LOCATION: 15242..15260
OTHER INFORMATION: 17-41-250.mis complement
NAME/KEY: primer_bind
LOCATION: 42199..42217
OTHER INFORMATION: 20-841-149.mis
NAME/KEY: primer_bind
LOCATION: 42219..42237
OTHER INFORMATION: 20-841-149.mis complement
NAME/KEY: primer_bind
LOCATION: 45423..45441
OTHER INFORMATION: 20-842-115.mis
NAME/KEY: primer_bind
LOCATION: 45443..45461
OTHER INFORMATION: 20-842-115.mis complement
NAME/KEY: primer_bind
LOCATION: 77039..77057
OTHER INFORMATION: 20-853-415.mis
NAME/KEY: primer_bind
LOCATION: 77059..77077
OTHER INFORMATION: 20-853-415.mis complement
NAME/KEY: misc_binding
LOCATION: 1227..1251
OTHER INFORMATION: 20-828-311.probe
NAME/KEY: misc_binding
LOCATION: 12335..12359
OTHER INFORMATION: 17-42-319.probe
NAME/KEY: misc_binding
LOCATION: 15229..15253
OTHER INFORMATION: 17-41-250.probe
NAME/KEY: misc_binding
LOCATION: 42206..42230
OTHER INFORMATION: 20-841-149.probe
NAME/KEY: misc_binding
LOCATION: 45430..45454
OTHER INFORMATION: 20-842-115.probe
NAME/KEY: misc_binding
LOCATION: 77046..77070
OTHER INFORMATION: 20-853-415.probe
US-09-751-877-1

Query Match
Best Local Similarity 85.6%; Score 15.4; DB 10; Length 81001;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
RESULT 6
US-09-878-574-5808
; Sequence 5808, Application US/09878574
; Patent No. US20020110548A1
; GENERAL INFORMATION:
```

```
APPLICANT: Byrum, Joseph R.
APPLICANT: La Rosa, Thomas J.
APPLICANT: Thompson, Michael D.
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
FILE REFERENCE: 38-21(15401)B
CURRENT APPLICATION NUMBER: US/09/878,574
PRIOR FILING DATE: 2001-12-21
PRIOR APPLICATION NUMBER: 09/333,535
NUMBER OF SEQ ID NOS: 15775
SEQ ID NO 5808
LENGTH: 171
TYPE: DNA
ORGANISM: glycine max
OTHER INFORMATION: Clone ID: 701097386H1
US-09-878-574-5808

Query Match
Best Local Similarity 83.3%; Score 15; DB 10; Length 171;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 CTCAGAGAGAGAGCA 15
DB 98 CTCAGAGAGAGAGCA 112
```

```
RESULT 7
US-09-563-817-31
; Sequence 31, Application US/09563817
; Patent No. US2002095031A1
; GENERAL INFORMATION:
; APPLICANT: Nehls, Michael C.
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: Polypeptides Encoded Thereby
; FILE REFERENCE: LEX-0021-USA
; CURRENT APPLICATION NUMBER: US/09/563,817
; PRIOR FILING DATE: 2001-06-11
; PRIOR APPLICATION NUMBER: US 60/132,343
; NUMBER OF SEQ ID NOS: 1008
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 258
; TYPE: DNA
; ORGANISM: homo sapiens
US-09-563-817-31
```

```
Query Match
Best Local Similarity 83.3%; Score 15; DB 10; Length 258;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 2 TCAGAGAGAGAGAG 16
DB 54 TCAGAGAGAGAGAG 68
```

```
RESULT 8
US-09-733-607-7
; Sequence 7, Application US/09733607
; Patent No. US20020042054A1
; GENERAL INFORMATION:
; APPLICANT: Tononi, Giulio
; APPLICANT: Cirelli, Chiara
; APPLICANT: Shaw, Paul J.
; APPLICANT: Greenspan, Ralph J.
; TITLE OF INVENTION: Vigilance Nucleic Acids and Related
; FILE REFERENCE: P-NI 4447
; CURRENT APPLICATION NUMBER: US/09/733,607
; CURRENT FILING DATE: 2001-09-10
```

```
;; PRIOR APPLICATION NUMBER: US 09/456,785
;; PRIOR FILING DATE: 1999-12-08
;; NUMBER OF SEQ ID NOS: 27
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 7
;; LENGTH: 267
;; TYPE: DNA
;; ORGANISM: Drosophila
US-09-733-607-7

Query Match      83.3%; Score 15; DB 10; Length 267;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      4 AGAGAGAGAGAGGCC 18
Db      146 AGAGAGAGAGAGGCC 160
      |||
      |||

RESULT 9
US-09-878-574-7091
; Sequence 7091, Application US/09878574
; Patent No. US20020110548A1
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Thompson, Michael D.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(15401)B
; CURRENT APPLICATION NUMBER: US/09/878,574
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/333,535
; PRIOR FILING DATE: 1999-06-14
; NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 7091
; LENGTH: 272
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: 701099415H1
US-09-878-574-7091

Query Match      83.3%; Score 15; DB 10; Length 272;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 TCAGAGAGAGAGGAG 16
Db      59 TCAGAGAGAGAGGAG 73
      |||
      |||

RESULT 10
US-09-878-574-11646/C
; Sequence 11646, Application US/09878574
; Patent No. US20020110548A1
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Thompson, Michael D.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(15401)B
; CURRENT APPLICATION NUMBER: US/09/878,574
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/333,535
; PRIOR FILING DATE: 1999-06-14
; NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 11646
; LENGTH: 283
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: 701064705H1
US-09-878-574-11646
```

```
Query Match      83.3%; Score 15; DB 10; Length 283;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 TCAGAGAGAGAGGAG 16
Db      22 TCAGAGAGAGAGGAG 8
      |||
      |||

RESULT 11
US-09-864-761-11469/C
; Sequence 11469, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Aecm1ca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 11469
; LENGTH: 481
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC006065.3
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.81
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.65
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.74
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.71
```

OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.82
US-09-864-761-11469

Query Match 83.3%; Score 15; DB 10; Length 481;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 AGAGAGAGAGAGGCC 18
DB 225 AGAGAGAGAGAGGCC 211

RESULT 12
US-09-924-401-19/c
; Sequence 19, Application US/09924401
; Patent No. US20020142957A1
; GENERAL INFORMATION:
; APPLICANT: Hepler, William T.
; APPLICANT: Jlang, Yuguu
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jlangchun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.550
; CURRENT APPLICATION NUMBER: US/09/924,401
; CURRENT FILING DATE: 2001-08-07
; NUMBER OF SEQ ID NOS: 85
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 547
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-924-401-19

Query Match 83.3%; Score 15; DB 10; Length 547;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TCAGAGAGAGAGAGGCC 16
DB 149 TCAGAGAGAGAGAGGCC 135

RESULT 13
US-09-960-253-74
; Sequence 74, Application US/09960253
; Patent No. US20020123619A1
; GENERAL INFORMATION:
; APPLICANT: Benson, Darin R.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Lodes, Michael J.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.556
; CURRENT APPLICATION NUMBER: US/09/960,253
; CURRENT FILING DATE: 2001-09-20
; NUMBER OF SEQ ID NOS: 187
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 74
; LENGTH: 5540
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-960-253-74

Query Match 83.3%; Score 15; DB 10; Length 5540;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TCAGAGAGAGAGAGGCC 16
DB 2157 TCAGAGAGAGAGAGGCC 2171

RESULT 14
US-09-867-701-8586
; Sequence 8586, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8586
; LENGTH: 183
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-867-701-8586

Query Match 82.2%; Score 14.8; DB 10; Length 183;
Best Local Similarity 88.9%; Pred. No. 2e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTCAGAGAGAGAGAGGCC 18
DB 112 CTCAGAGAGAGAGAGGCC 129

RESULT 15
US-09-960-352-12042
; Sequence 12042, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Mengding
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(1029)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 12042
; LENGTH: 236
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 51-LIB34-068-Q1-E1-E4
US-09-960-352-12042

Query Match 82.2%; Score 14.8; DB 10; Length 236;
Best Local Similarity 88.9%; Pred. No. 2.1e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTCAGAGAGAGAGAGGCC 18
DB 206 CTCAGAGAGAGAGAGGCC 223

Search completed: November 9, 2002, 14:07:31
Job time : 16.4007 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 9, 2002, 09:08:56 ; Search time 5.16041 Seconds

(without alignments)
7855.185 Million cell updates/sec

Title: US-09-902-772-1_COPY_645_662

Perfect score: 18
Sequence: 1 ctcagagagagagagc 18

Scoring table: IDENTITY_NDC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_101002:*

1: /SID2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:*
2: /SID2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:*
3: /SID2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:*
4: /SID2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:*
5: /SID2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:*
6: /SID2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:*
7: /SID2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:*
8: /SID2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:*
9: /SID2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:*
10: /SID2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:*
11: /SID2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:*
12: /SID2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:*
13: /SID2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:*
14: /SID2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:*
15: /SID2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:*
16: /SID2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:*
17: /SID2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:*
18: /SID2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:*
19: /SID2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:*
20: /SID2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:*
21: /SID2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:*
22: /SID2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:*
23: /SID2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:*
24: /SID2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18	100.0	1447	20	AA26551
2	16.4	91.1	1422	23	AB197996
3	16.4	91.1	1522	21	AA21022
4	16.4	91.1	1522	21	AAA34900
5	16.4	91.1	1956	15	AAQ73587
6	16.4	91.1	2100	22	AA502501
7	16.4	91.1	2100	22	AA502502
8	16.4	91.1	2100	22	AA502503
9	16.4	91.1	2100	22	AA502504

10	16.4	91.1	2100	22	AA502505	Human gene for m1
11	16.4	91.1	2100	22	AA502506	Human gene for m1
12	16.4	91.1	2100	22	AA502508	Human gene for m1
13	16.4	91.1	2880	10	AA002068	Human muscarinic a
14	16.4	91.1	2880	10	AA92068	Human muscarinic a
15	16.4	91.1	3150	24	AB92262	Prostate cancer as
16	16.4	91.1	4132	21	AA36650	Nucleotide sequenc
17	16.4	91.1	4204	24	AA74417	Tissue remodeling
18	16.4	91.1	4204	24	ABR84512	Human cDNA differe
19	16.4	91.1	4204	24	ABR84800	Human benign prost
20	16.4	91.1	4231	22	AA544672	Human full-length
21	16.4	91.1	4251	24	AA594851	Human DNA sequence
22	16.4	91.1	6707	21	AA21026	Human low adenosin
23	16.4	91.1	6707	21	AAA34904	Human adenosine re
24	16.4	91.1	611590	21	AA22303	Arabidopsis thalia
25	16	88.9	39678	22	AAK84996	Human immune/haema
26	15.4	85.6	65	24	ABN30396	Rat spliced trans
27	15.4	85.6	170	22	ABN75834	Human foetal liver
28	15.4	85.6	170	22	ABA40401	Probe #18867 for g
29	15.4	85.6	170	22	AAK24515	Human brain expres
30	15.4	85.6	170	22	AAK50504	Human bone marrow
31	15.4	85.6	170	22	AA127517	Probe #17450 for g
32	15.4	85.6	170	22	AA156489	Probe #25175 used
33	15.4	85.6	170	24	AB523978	Human genome-deriv
34	15.4	85.6	347	22	AAK81511	Human immune/haema
35	15.4	85.6	347	22	AAK81512	Human immune/haema
36	15.4	85.6	371	22	AA182088	Human polynucleoti
37	15.4	85.6	475	23	ABV48538	Human prostate exp
38	15.4	85.6	506	24	ABK63037	Rat sequence diffe
39	15.4	85.6	583	22	ABA63430	Human foetal liver
40	15.4	85.6	583	22	ABA30629	Probe #9095 for ge
41	15.4	85.6	583	22	AAK11963	Human brain expres
42	15.4	85.6	583	22	AAK37668	Human bone marrow
43	15.4	85.6	583	22	AA118425	Probe #8358 for ge
44	15.4	85.6	583	22	AA143541	Probe #12227 used
45	15.4	85.6	583	24	AB511656	Human genome-deriv

ALIGNMENTS

RESULT 1	AA26551	standard; DNA; 1447 BP.
ID	AA26551	
XX	AA26551:	
XX	14-JUN-1999	(first entry)
XX	DE	DNA encoding chicken C-11 protein.
XX	XX	Chicken: C-11 protein; cell calcification inhibiting activity;
XX	KW	cell calcification inhibiting agent; c-ery protein; arthritis deformans;
XX	KW	ossification; spinal column ligament; ss.
XX	XX	
OS	Gallus sp.	
XX	XX	
FH	Key	Location/Qualifiers
FT	CDS	63..1418
FT		/*tag- a
XX	XX	
PN	JP11075871-A.	
XX	XX	
PD	23-MAR-1999.	
XX	XX	
PF	29-MAY-1998;	98JP-0166076.
XX	XX	
PR	20-JUN-1997;	97US-0050297.
XX	18-JUN-1997;	97US-0878177.
XX	XX	
PA	(CHUS) CHUGAI PHARM CO LTD.	
XX	(UYPE-) UNIV PENNSYLVANIA.	
XX	XX	

DR WPI; 1999-257708/22.
DR P-PSDB; AAY01520.

An active protein for inhibiting cell calcification - useful for measuring the calcification of a cell, for diagnosing arthritis deformans or ossification of spinal column ligament

Pt

xx

ps Disclosure; Page 7-8; 15pp; Japanese.

The present sequence encodes a chicken C-11 protein which has cell calcification inhibiting activity. The specification also describes a cell calcification inhibiting agent containing c-reg protein (AAY01521). The proteins are used for measuring the calcification of a cell, for diagnosing arthritis deformans or ossification of spinal column ligament.

CC

cc

XX

SQ Sequence 1447 BP; 440 A; 374 C; 317 G; 316 T; 0 other;

Query Match Best Local Similarity 100.0%; Score 18; DB 20; Length 1447;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Gy 1 CTCAGGAGAGAGACC 18
|||||||
Db 645 CTCAGGAGAGAGACC 662

RESULT 2
ABI97996
ID ABI97996 standard; CDNA; 1422 BP.

AC ABI97996;
XX
Dt 18-FEB-2002 (first entry)
XX
Xx Non-endogenous human GPCR CDNA, SMO ID NO: 512.
DE
XX
KW Human; G protein-coupled receptor; GPCR; non-endogenous; mutant; constitutively activated GPCR; agonist; disease; ss.
KM
OS Homo sapiens.
SS Synthetic.
PN WO200177172-A2.
PD 18-OCT-2001.
PF 05-APR-2001; 2001WO-US11098.
PR 07-APR-2000; 2000US-19574P.
PA (AREN-) ARENA PHARM INC.
Pl Lehmann-Bruinsma K; Llaw CW, Lin I;
WI; 2001-648759/74.
PS PSDB; ABB56360.

Identifying agonists of G protein-coupled receptors (GPCRs) for use in disease treatment, comprises contacting candidate compounds with versions of GPCRS -

Example 2: Page 312; 394pp; English.

The invention relates to G protein-coupled receptors (GPCRs) for which the endogenous ligand has been identified. Non-endogenous constitutively activated versions of known GPCRs are used in the invention for the direct identification of candidate compounds as receptor agonists, inverse agonists or partial agonists. Such agonists are useful as therapeutic agents for diseases or disorders associated with GPCRs. The present sequence encodes a non-endogenous version of a known human GPCR.

Seq	Sequence	1422 BP	286 A	482 C	388 G	266 T	0 other
Query Match		91.1%	Score 16.4	DB 23	Length 1422		
Best Local Similarity		94.4%	Pred. No. 2.6e+02				
Matches	17	Conservative	0	Mismatches	1	Indels	0
QY	1 CTCACAGCAGCAGCAGC 18						
Db	870 CTCACAGCAGCAGCAGC 887						
RESULT 3							
AAEF21022							
ID	AAEF21022 standard; DNA; 1522 BP.						
XX							
AC	AAEF21022;						
XX							
DT	14-MAR-2001 (first entry)						
XX							
DE	Human low adenosine antisense oligonucleotide related sequence #2589.						
XX							
KW	Low adenosine antisense oligonucleotide; phosphorothioate; allergy;						
KW	human; airway disorder; bronchoconstriction; lung inflammation;						
KW	surfactant depletion; respiratory; bronchodilator; antiinflammatory;						
KW	immunosuppressive; antiasthmatic; analgesic; hypotensive; cyostatic;						
KW	respiratory obstruction; pulmonary obstruction; impeded respiration;						
KW	surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;						
KW	respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;						
KW	pulmonary hypertension; emphysema; pulmonary transplantation rejection;						
KW	chronic obstructive pulmonary disease; pulmonary infection; bronchitis;						
KW	cancer; ss.						
XX							
OS	Homo sapiens.						
XX							
PN	WO200062736-A2.						
XX							
PD	26-OCT-2000.						
XX							
PF	24-MAR-2000; 2000WO-US08020.						
XX							
PR	06-APR-1999; 99US-0127958.						
XX							
PA	(UYEC-) UNIV EAST CAROLINA.						
XX	(NYCE/) NYCE J W.						
XX							
PI	Nyce JW;						
XX							
DR	WPI; 2000-679539/66.						
XX							
PT	Low adenosine (A) content antisense oligonucleotides which do not						
PT	trigger adenosine receptors during metabolism, useful e.g. for treating						
PT	cancers and respiratory obstructions -						
XX							
PS	Disclosure; Page 832-833; 15922P; English.						
XX							
CC	The present invention describes low adenosine (A) content antisense						
CC	oligonucleotides and compositions (I) comprising them. In the antisense						
CC	oligonucleotides the A is replaced by a "universal" or alternative base.						
CC	(I) can have respiratory, bronchodilator, antiinflammatory, analgesic,						
CC	immunosuppressive, antiasthmatic, hypotensive and cyostatic activities.						
CC	The antisense oligonucleotides and (I) can be used to down-regulate the						
CC	expression and or activity of target polypeptides associated with						
CC	lung/respiratory disorders and malignancies, such as stimulating and						
CC	activating peptide factors and transmitters, transcription factors,						
CC	immunoglobulins and antibodies, antibody receptors, cytokines and						
CC	chemokines, endogenously produced specific and non-specific enzymes,						
CC	binding proteins, adhesion molecules and their receptors, cytokine and						
CC	chemokine receptors, adenosine receptors, bradykinin receptors, central						
CC	nervous system (CNS) and peripheral nervous and non-nervous system						
CC	receptors, CNS and peripheral nervous and non-nervous system peptide						
CC	transmitters, defensins, growth factors, vasoactive peptides and						
CC	receptors, binding proteins and malignancy associated proteins. The						
CC	antisense oligonucleotides may be used in this way to treat disorders						

CC including respiratory obstruction (especially pulmonary obstruction
 CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
 CC and/or surfactant hypoproduction which are associated with a disease or
 CC condition selected from pulmonary vasoconstriction, inflammation,
 CC allergies, asthma, impeded respiration, respiratory distress syndrome
 CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
 CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
 CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
 CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
 CC fragments and antisense oligonucleotides used in the exemplification of
 CC the present invention.

SQ Sequence 1522 BP; 301 A; 520 C; 413 G; 288 T; 0 other;

Query Match 91.1%; Score 16.4; DB 21; Length 1522;
 Best Local Similarity 94.4%; Pred. No. 2.6e+02;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 CTCGAGAGGAGGAGGCC 18
 ||||||| |||||||
 Db 970 CTCGAGGAGGAGGAGGCC 987

RESULT 4
 AAA34900
 ID AAA34900 standard; DNA: 1522 BP.
 AC AAA34900;
 XX
 XX 28-JUL-2000 (first entry)
 DT
 XX
 XX Human adenosine receptor related polynucleotide SEQ ID NO:3589.
 DE
 XX
 XX Human; adenosine receptor; low adenosine antisense oligonucleotide;
 KM phosphorothioate; impeded respiration; inflammation; allergy;
 KM allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
 KM antiallergic; antiasthmatic; cytosstatic; analgesic; impaired airway;
 KM lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
 KM respiratory distress syndrome; pain; cystic fibrosis; emphysema;
 KM pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
 KM cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.
 KM
 XX
 XX Homo sapiens.
 OS
 XX
 XX WO200009525-A2.
 PN
 XX
 XX 24-FEB-2000.
 PD
 XX
 XX 03-AUG-1999; 99WO-US17712.
 PF
 XX
 XX 03-AUG-1998; 98US-0095212.
 PR
 XX
 XX (UYEC-) UNIV EAST CAROLINA.
 PA
 XX
 XX Nyce JW;
 PI
 XX
 XX WPI; 2000-205971/18.
 DR
 XX
 XX New antisense oligonucleotides useful for treating e.g. pulmonary
 PT vasoconstriction, inflammation, allergies, asthma, hypertension,
 PT bronchitis, emphysema, respiratory distress syndrome, ischemia or
 PT cancers
 PT
 XX
 XX Disclosure; Page 762; 1343pp; English.
 PS
 XX
 XX The present invention describes a new composition comprising an
 CC antisense oligonucleotide (ON) with low adenosine (up to 15%), which
 CC targets nucleic acids involved in bronchoconstriction, allergies, and/or
 CC inflammation. The ON can have antiinflammatory, antiallergic,
 CC antiasthmatic, cytosstatic and analgesic activities. The compositions
 CC are useful for the treatment of diseases associated with inflammation,
 CC impaired airways, including lung disease and diseases whose secondary
 CC effects afflict the lungs of a subject. They can be used for treating

CC e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,
 CC asthma, impeded respiration, respiratory distress syndrome, pain, cystic
 CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive
 CC pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,
 CC carcinomas, and cancers which may metastasise to the lungs, including
 CC breast and prostate cancer. The reduction of the adenosine content of
 CC the ONs reduces side effects. The A-containing ONs break down with the
 CC release of deoxyadenosine which activates adenosine receptors causing
 CC bronchoconstriction and inflammation. AAA32313 to AAA5312 represent the
 CC nucleotide sequences given in the sequence listing from the present
 CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last
 CC 185 sequences are also called SEQ ID NO:1 to 185, but the sequences
 CC differ from the previously named sequences. SEQ ID NO:11 to 1680
 CC (AAA32323 to AAA33992) are specifically claimed ONs from the present
 CC invention. N.B. Sequences given in the disclosure of the present
 CC invention do not match up with their corresponding SEQ ID NO: sequences
 CC given in the sequence listing.

SQ Sequence 1522 BP; 301 A; 520 C; 413 G; 288 T; 0 other;

Query Match 91.1%; Score 16.4; DB 21; Length 1522;
 Best Local Similarity 94.4%; Pred. No. 2.6e+02;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 CTCGAGAGGAGGAGGCC 18
 ||||||| |||||||
 Db 970 CTCGAGGAGGAGGAGGCC 987

RESULT 5
 AA073587
 ID AA073587 standard; DNA: 1956 BP.
 AC AA073587;
 XX
 XX 25-JUN-1995 (first entry)
 DT
 XX
 XX Fragment of the muscarinic acetylcholine receptor.
 DE
 XX
 XX AChR; expression; ss.
 KM
 KM
 XX
 XX Homo sapiens.
 OS
 XX
 XX Key Location/Qualifiers
 FH CDS 376..1815
 FT /*tag= a
 FT
 XX
 XX WO9421789-A.
 PN
 XX
 XX 29-SEP-1994.
 PD
 XX
 XX 28-FEB-1994; 94WO-US02388.
 PF
 XX
 XX 25-MAR-1993; 93US-0038662.
 PR
 XX
 XX (REGC) UNIV CALIFORNIA.
 PA
 XX
 XX Betlach MC, Turner GJ;
 PI
 XX
 XX WPI; 1994-317010/39.
 DR
 XX
 XX P-PsDB; AAR60695.
 PT
 XX
 XX Expression of heterologous proteins in halo-bacteria - using
 PT regulatory and stop sequences from halo-bacteria, pref. the
 PT bacterio-rhodopsin gene.
 PT
 XX
 XX Disclosure; Fig 6; 118pp; English.
 PS
 XX
 XX The sequence is that of the human muscarinic acetylcholine receptor
 CC gene fragment. The gene is used to exemplify
 CC a new expression vector for producing heterologous polypeptides in a
 CC halo-bacterial host.
 CC
 CC See also AA073586-92.

XX	13-OCT-1999;	99US-0159269.
XX		
PA	(GENA-) GENAISSANCE PHARM INC.	
XX		
PI	Choi JY, Denton RR, Nandabalan K, Stephens JC;	
XX		
XX	WPI: 2001-282046/29.	
DR	P-PSDB; AAU01294.	
XX		
PT	New variants of the m1 muscarinic acetylcholine receptor gene, useful	
PT	to find treatment for Alzheimer's and dementia, have single nucleotide	
PT	variations at one or more of five polymorphic sites -	
XX		
PS	Claim 26; Page - : 52pp; English.	
XX		
CC	The sequence represents the Human gene encoding the m1 muscarinic	
CC	acetylcholine receptor, CHRM1, haplotype 1. CHRM1 is one subtype of a	
CC	family of 5 genetically distinct muscarinic acetylcholine receptors,	
CC	mAChR, that play important roles in higher brain function such as	
CC	learning and memory. The protein is a possible drug target for treatments	
CC	for Alzheimer's disease and dementia with Lewy bodies (DLB). The gene,	
CC	polypeptide, haplotypes and antibodies raised against the protein are	
CC	useful for diagnosing and developing treatments for diseases associated	
CC	with the abnormal expression of the gene or activity of the protein, e.g.	
CC	Alzheimer's disease and dementia with Lewy bodies.	
CC	Note: The present sequence is not shown in the specification but is	
CC	derived from the CHRM1 sequence given in AAS02501 and information	
CC	contained in table 4 (page 31) of the specification.	
XX		
XX	Sequence 2100 BP; 460 A; 661 C; 570 G; 409 T; 0 other;	
QY	Query Match	91.1%; Score 16.4; DB 22; Length 2100;
	Best Local Similarity	94.4%; Pred. No.2.6e+02;
	Matches 17; Conservative	0; Mismatches 1; Indels 0; Gaps 0;
DB	1 CTCAGAGAGAGAGAGCC 18	
	1320 CTCAGAGGAGAGAGGCC 1337	
RESULT 8		
AAS02503		
ID	AAS02503 standard; DNA: 2100 BP.	
XX		
AC	AAS02503;	
XX		
DT	29-AUG-2001 (first entry)	
XX		
DE	Human gene for m1 muscarinic acetylcholine receptor, haplotype 2.	
XX		
KW	Human; m1 acetylcholine receptor; CHRM1; immunogen; antibody;	
KW	Alzheimer's disease; dementia with Lewy bodies; DLB; haplotype 2; ds.	
XX		
OS	Homo sapiens.	
XX		
FH	Key	Location/Qualifiers
FT	intron	1..450
FT		/*tag= a
FT		/number= 1
FT	exon	451..1833
FT		/*tag= b
FT		/number= 1
FT	CDS	451..1833
FT		/*tag= c
FT		/product= "CHRM1 protein"
FT	variation	/note= "this CDS is specifically claimed"
FT		replace (717,C)
FT		/*tag= d
FT		/label= "single_nucleotide_polymorphism"
FT		replace (1803,C)
FT		/*tag= e
FT	variation	/label= "single_nucleotide_polymorphism"
FT		

FT	Intron	1834..2100
FT		/*tag= f
FT		/number= 2
PN	MO200127312-A2.	
XX	19-APR-2001.	
PD		
XX	12-OCT-2000; 2000WO-US28211.	
XX	13-OCT-1999; 99US-0159269.	
PR		
XX	(GENA-) GENAISSANCE PHARM INC.	
PA		
PI	Choi JY, Denton RR, Nandabalan K, Stephens JC;	
XX		
DR	WPI: 2001-282046/29.	
DR	P-PSDB: AAU01293.	
XX		
PT	New variants of the m1 muscarinic acetylcholine receptor gene, useful	
PT	to find treatment for Alzheimer's and dementia, have single nucleotide	
XX	variations at one or more of five polymorphic sites	
PS	Claim 26; Page - : 52pp; English.	
XX		
CC	The sequence represents the Human gene encoding the m1 muscarinic	
CC	acetylcholine receptor, CHMRL, haplotype 2. CHMRL is one subtype of a	
CC	family of 5 genetically distinct muscarinic acetylcholine receptors,	
CC	MACHR, that play important roles in higher brain function such as	
CC	learning and memory. The protein is a possible drug target for treatments	
CC	for Alzheimer's disease and dementia with Lewy bodies (DLB). The gene,	
CC	polypeptide, haplotypes and antibodies raised against the protein are	
CC	useful for diagnosing and developing treatments for diseases associated	
CC	with the abnormal expression of the gene or activity of the protein, e.g.	
CC	Alzheimer's disease and dementia with Lewy bodies.	
CC	Note: The present sequence is not shown in the specification but is	
CC	derived from the CHMRL sequence given in AAS02501 and information	
CC	contained in table 4 (page 31) of the specification.	
XX		
SEQ	Sequence 2100 BP; 459 A; 661 C; 571 G; 409 T; 0 other;	
	Query Match	91.1%; Score 16.4; DB 22; Length 2100;
	Best Local Similarity	94.4%; Pred. No. 2.6e+02;
	Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
QY	1 CTCAGAGAGAGGAGGCC 18	
Db	1320 CTCAGAGGGAGAGGAGCC 1337	
RESULT 9		
AAS02504		
ID	AAS02504 standard; DNA; 2100 BP.	
XX		
AC	AAS02504;	
XX		
DT	29-AUG-2001 (first entry)	
XX		
DE	Human gene for m1 muscarinic acetylcholine receptor, haplotype 3.	
XX		
KW	Human; m1 acetylcholine receptor; CHMRL; immunogen; antibody;	
XX	Alzheimer's disease; dementia with Lewy bodies; DLB; haplotype 3; ds.	
OS	Homo sapiens.	
XX		
Key	Location/Qualifiers	
FT	1..450	
FT	intron	/*tag= a
FT		/number= 1
FT	exon	451..1833
FT		/*tag= b
FT		/number= 1
FT	CDS	451..1833

```
FT FT      /*tag= C
FT FT      /product= "CHRM1 protein"
FT FT      /note= "This CDS is specifically claimed"
FT FT      replace (717,C)
FT FT      /*tag= d
FT FT      /label= "Single-nucleotide-polymorphism"
FT FT      replace (1233,C)
FT FT      /*tag= e
FT FT      /label= "Single-nucleotide-polymorphism"
FT FT      replace (1803,C)
FT FT      /*tag= f
FT FT      /label= "Single-nucleotide-polymorphism"
FT FT      1834..2100
FT FT      /*tag= 9
FT FT      /number= 2
FT FT
FT FT      WO200127312-A2.
FT FT
FT FT      19-APR-2001.
FT FT
FT FT      12-OCT-2000; 2000WO-US28211.
FT FT
FT FT      13-OCT-1999; 99US-0159269.
FT FT
FT FT      (GENA-) GENAISSANCE PHARM INC.
FT FT
FT FT      Choi JY, Denton RR, Nandabalan K, Stephens JC;
FT FT      MPI; 2001-282046/29.
FT FT      P-PSDB; AAU01293.
FT FT
FT FT      New variants of the m1 muscarinic acetylcholine receptor gene, useful
FT FT      to find treatment for Alzheimer's and dementia, have single nucleotide
FT FT      variations at one or more of five polymorphic sites -
FT FT
FT FT      Claim 26; Page - : 52pp; English.
FT FT
FT FT      The sequence represents the Human gene encoding the m1 muscarinic
FT FT      acetylcholine receptor, CHMR1, haplotype 3. CHMR1 is one subtype of a
FT FT      family of 5 genetically distinct muscarinic acetylcholine receptors,
FT FT      mACHR, that play important roles in higher brain function such as
FT FT      learning and memory. The protein is a possible drug target for treatments
FT FT      for Alzheimer's disease and dementia with Lewy bodies (DLB). The gene,
FT FT      polypeptide, haplotypes and antibodies raised against the protein are
FT FT      useful for diagnosing and developing treatments for diseases associated
FT FT      with the abnormal expression of the gene or activity of the protein, e.g.
FT FT      CC Alzheimer's disease and dementia with Lewy bodies.
FT FT      CC Note: The present sequence is not shown in the specification but is
FT FT      derived from the CHMR1 sequence given in AAS02501 and information
FT FT      contained in table 4 (page 31) of the specification.
FT FT
FT FT      Sequence 2100 BP; 459 A; 660 C; 571 G; 410 T; 0 other;
FT FT
FT FT      Query Match      91.1%; Score 16.4; DB 22; Length 2100;
FT FT      Best Local Similarity 94.4%; Pred. No. 2.6e-02;
FT FT      Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
FT FT
FT FT      OY      1 CTCAGAGAGAGAGAGACC 18
FT FT      ||||||| |||||||
FT FT      DB      1320 CTCAGAGAGAGAGAGACC 1337
FT FT
FT FT      RESULT 10
FT FT      AAS02505
FT FT      ID      AAS02505 standard; DNA; 2100 BP.
FT FT      XX
FT FT      AC      AAS02505;
FT FT
FT FT      29-AUG-2001 (first entry)
FT FT
FT FT      Human gene for m1 muscarinic acetylcholine receptor, haplotype 4.
FT FT      DE      Human; m1 acetylcholine receptor; CHMR1; immunogen; antibody;
FT FT      XX
FT FT      KW
```

```
KW Alzheimer's disease; dementia with Lewy bodies; DLB; haplotype 4; ds.
XX Homo sapiens.
OS
XX
XX Key
FH      Location/Qualifiers
FT      Intron
FT      1..450
FT      /*tag= a
FT      /number= 1
FT      exon
FT      451..1833
FT      /*tag= b
FT      /number= 1
FT      CDS
FT      451..1833
FT      /*tag= C
FT      /product= "CHRM1 protein"
FT      /note= "This CDS is specifically claimed"
FT      replace (1199,C)
FT      /*tag= d
FT      /label= "Single-nucleotide-polymorphism"
FT      1834..2100
FT      /*tag= e
FT      /number= 2
FT FT
FT FT      WO200127312-A2.
FT FT
FT FT      19-APR-2001.
FT FT
FT FT      12-OCT-2000; 2000WO-US28211.
FT FT
FT FT      13-OCT-1999; 99US-0159269.
FT FT
FT FT      (GENA-) GENAISSANCE PHARM INC.
FT FT
FT FT      Choi JY, Denton RR, Nandabalan K, Stephens JC;
FT FT      MPI; 2001-282046/29.
FT FT      P-PSDB; AAU01294.
FT FT
FT FT      New variants of the m1 muscarinic acetylcholine receptor gene, useful
FT FT      to find treatment for Alzheimer's and dementia, have single nucleotide
FT FT      variations at one or more of five polymorphic sites -
FT FT
FT FT      Claim 26; Page - : 52pp; English.
FT FT
FT FT      The sequence represents the Human gene encoding the m1 muscarinic
FT FT      acetylcholine receptor, CHMR1, haplotype 4. CHMR1 is one subtype of a
FT FT      family of 5 genetically distinct muscarinic acetylcholine receptors,
FT FT      mACHR, that play important roles in higher brain function such as
FT FT      learning and memory. The protein is a possible drug target for treatments
FT FT      for Alzheimer's disease and dementia with Lewy bodies (DLB). The gene,
FT FT      polypeptide, haplotypes and antibodies raised against the protein are
FT FT      useful for diagnosing and developing treatments for diseases associated
FT FT      with the abnormal expression of the gene or activity of the protein, e.g.
FT FT      CC Alzheimer's disease and dementia with Lewy bodies.
FT FT      CC Note: The present sequence is not shown in the specification but is
FT FT      derived from the CHMR1 sequence given in AAS02501 and information
FT FT      contained in table 4 (page 31) of the specification.
FT FT
FT FT      Sequence 2100 BP; 459 A; 663 C; 570 G; 408 T; 0 other;
FT FT
FT FT      Query Match      91.1%; Score 16.4; DB 22; Length 2100;
FT FT      Best Local Similarity 94.4%; Pred. No. 2.6e-02;
FT FT      Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
FT FT
FT FT      OY      1 CTCAGAGAGAGAGAGACC 18
FT FT      ||||||| |||||||
FT FT      DB      1320 CTCAGAGAGAGAGAGACC 1337
FT FT
FT FT      RESULT 11
FT FT      AAS02506
FT FT      ID      AAS02506 standard; DNA; 2100 BP.
FT FT      XX
FT FT      AC      AAS02506;
```

```
XX 29-AUG-2001 (first entry)
DT
XX Human gene for m1 muscarinic acetylcholine receptor, haplotype 5.
DE
XX Human: m1 acetylcholine receptor; CHRM1; immunogen; antibody;
KM Alzheimer's disease; dementia with Lewy bodies; DLB; haplotype 5; ds.
XX
XX Homo sapiens.
OS
XX
XX Key
FH Location/Qualifiers
FT intron
FT 1..450
FT /*tag= a
FT /number= 1
FT exon
FT 451..1833
FT /*tag= b
FT /number= 1
FT CDS
FT 451..1833
FT /*tag= c
FT /product= "CHRM1 protein"
FT /note= "This CDS is specifically claimed"
FT variation
FT replace (1494,G)
FT /*tag= d
FT /label= "Single-nucleotide-polymorphism"
FT intron
FT 1834..2100
FT /*tag= e
FT /number= 2
FT
FT WO200127312-A2.
XX
XX 19-APR-2001.
XX
XX 12-OCT-2000; 2000WO-US28211.
XX
XX 13-OCT-1999; 99US-0159269.
XX
XX (GENA-) GENMAISSANCE PHARM INC.
XX
XX ChOI JY, Denton RR, Nandabalan K, Stephens JC;
PI
XX WPI: 2001-282046/29.
XX
XX P-PSDB; AAU01293.
XX
XX New variants of the m1 muscarinic acetylcholine receptor gene, useful
XX to find treatment for Alzheimer's and dementia, have single nucleotide
XX variations at one or more of five polymorphic sites -
XX
XX Claim 26; Page - : 52pp; English.
XX
XX The sequence represents the Human gene encoding the m1 muscarinic
XX acetylcholine receptor, CHRM1, haplotype 5. CHRM1 is one subtype of a
XX family of 5 genetically distinct muscarinic acetylcholine receptors,
XX mAChR, that play important roles in higher brain function such as
XX learning and memory. The protein is a possible drug target for treatments
XX for Alzheimer's disease and dementia with Lewy bodies (DLB). The gene,
XX polypeptide, haplotypes and antibodies raised against the protein are
XX useful for diagnosing and developing treatments for diseases associated
XX with the abnormal expression of the gene or activity of the protein, e.g.
XX Alzheimer's disease and dementia with Lewy bodies.
XX Note: The present sequence is not shown in the specification but is
XX derived from the CHRM1 sequence given in AAS02501 and Information
XX contained in table 4 (page 31) of the specification.
XX
XX Sequence 2100 BP; 459 A; 663 C; 570 G; 408 T; 0 other;
XX
XX Query Match 91.1%; Score 16.4; DB 22; Length 2100;
XX Best Local Similarity 94.4%; Pred. No. 2.6e+02;
XX Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX 1 CTCAGAGAGAGAGAGCC 18
XX ||||| |||||
Db 1320 CTCAGAGGAGAGAGGCC 1337
```

```
RESULT 12
AAS02508
ID AAS02508 standard; DNA; 2100 BP.
XX
XX AAS02508;
AC
XX 29-AUG-2001 (first entry)
DT
XX Human gene for m1 muscarinic acetylcholine receptor, haplotype 7.
XX
XX Human: m1 acetylcholine receptor; CHRM1; immunogen; antibody;
KM Alzheimer's disease; dementia with Lewy bodies; DLB; haplotype 7; ds.
XX
XX Homo sapiens.
OS
XX
XX Key
FH Location/Qualifiers
FT intron
FT 1..450
FT /*tag= a
FT /number= 1
FT exon
FT 451..1833
FT /*tag= b
FT /number= 1
FT CDS
FT 451..1833
FT /*tag= c
FT /product= "CHRM1 protein"
FT /note= "This CDS is specifically claimed"
FT variation
FT replace (1803,C)
FT /*tag= d
FT /label= "Single-nucleotide-polymorphism"
FT intron
FT 1834..2100
FT /*tag= e
FT /number= 2
FT
FT WO200127312-A2.
XX
XX 19-APR-2001.
XX
XX 12-OCT-2000; 2000WO-US28211.
XX
XX 13-OCT-1999; 99US-0159269.
XX
XX (GENA-) GENMAISSANCE PHARM INC.
XX
XX ChOI JY, Denton RR, Nandabalan K, Stephens JC;
PI
XX WPI: 2001-282046/29.
XX
XX P-PSDB; AAS01293.
XX
XX New variants of the m1 muscarinic acetylcholine receptor gene, useful
XX to find treatment for Alzheimer's and dementia, have single nucleotide
XX variations at one or more of five polymorphic sites -
XX
XX Claim 26; Page - : 52pp; English.
XX
XX The sequence represents the Human gene encoding the m1 muscarinic
XX acetylcholine receptor, CHRM1, haplotype 7. CHRM1 is one subtype of a
XX family of 5 genetically distinct muscarinic acetylcholine receptors,
XX mAChR, that play important roles in higher brain function such as
XX learning and memory. The protein is a possible drug target for treatments
XX for Alzheimer's disease and dementia with Lewy bodies (DLB). The gene,
XX polypeptide, haplotypes and antibodies raised against the protein are
XX useful for diagnosing and developing treatments for diseases associated
XX with the abnormal expression of the gene or activity of the protein, e.g.
XX Alzheimer's disease and dementia with Lewy bodies.
XX Note: The present sequence is not shown in the specification but is
XX derived from the CHRM1 sequence given in AAS02501 and Information
XX contained in table 4 (page 31) of the specification.
XX
XX Sequence 2100 BP; 458 A; 662 C; 571 G; 409 T; 0 other;
XX
XX Query Match 91.1%; Score 16.4; DB 22; Length 2100;
XX Best Local Similarity 94.4%; Pred. No. 2.6e+02;
```

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Qy 1 CTCGAGAGAGAGAGACC 18
 ||||| |||||
 Db 1320 CTCGAGAGAGAGAGACC 1337

RESULT 13
 AAQ02068

ID AAQ02068 standard; DNA; 2880 BP.

XX
 AC AAQ02068;

DT 17-DEC-2001 (updated)
 DT 18-FEB-1999 (first entry)

DE Human muscarinic acetylcholine m1 receptor gene.

XX Muscarinic acetylcholine receptor; m1; drug screening; probes; ss.

XX Homo sapiens.

XX Key Location/Qualifiers
 FH cds 361..2740
 FT /tag= a
 FT /product=MAR subtype m1

FT misc_signal
 FT 282
 FT /tag= b
 FT /label=splice acceptor site
 FT /note=defines 5' end of exon
 FT 2780
 FT /tag= c

FT polyA_site
 FT 2780
 FT /tag= c

XX USN7241971-N.

XX 14-MAR-1989.

XX PD 08-SEP-1988; 88US-0241971.

XX PF 08-SEP-1988; 88US-0241971.

XX PR 08-SEP-1988; 88US-0241971.

XX PA (USSH) NAT INST OF HEALTH.

XX WPI; 1989-165452/22.
 DR P-PSDB; AAR06202.

XX Cloned genes for muscarinic acetylcholine receptors -
 PT for drug screening and diagnostic use.

XX PS Disclosure; ; PP; English.

XX The sequence was indexed from the best available specification copy;
 CC the unidentifiable bases were annotated as 'N'.

CC The sequence may be useful for synthesis of hybridisation probes for
 CC diagnostic use. The genes are cloned by screening a rat cerebral cortex

CC CDNA library with a probe based on nucleotides 170-225 of the non-coding
 CC strand of porcine brain (m1) cDNA (modified at positions 5, 38 and 53);

CC identifying cDNA clones by hybridisation with BamHI or XhoI digests of
 CC plasmid DNA from 12 cultures contg. up to 500000 independent clones;

CC recluturing until a pool contg. less than 5000 indep. clones with a
 CC single hybridisation band are identified, and isolating individual

CC clones by colony hybridisation.

CC stable cell lines are produced by transfecting Chinese hamster ovary
 CC cells (CHO-K1) with various pCD vectors contg. the gene inserts.

CC See also AAQ02068-Q020671 and AAQ00086.

CC (Note: Revised entry submitted to correct the patent number format of
 CC US Government-owned NTIS applications to prevent clashes with ongoing US

CC granted patent numbers. For further information please visit the Derwent
 CC web site at www.derwent.com/dwpl/updates/ntis.us.html.)

XX Sequence 2880 BP; 618 A; 921 C; 763 G; 550 T; 28 others; 0 other;

Query Match

91.1%; Score 16.4; DB 10; length 2880;

Best Local Similarity 94.4%; Pred. No. 2.7e+02;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Qy 1 CTCGAGAGAGAGAGACC 18
 ||||| |||||
 Db 1230 CTCGAGAGAGAGAGACC 1247

RESULT 14

AAAN92068

ID AAAN92068 standard; DNA; 2880 BP.

XX
 AC AAAN92068;

DT 17-DEC-2001 (updated)
 DT 27-SEP-1989 (first entry)

DE Human muscarinic acetylcholine m1 receptor gene.

XX Muscarinic acetylcholine receptor; m1; drug screening; probes; ss

XX Homo sapiens.

XX Key Location/Qualifiers
 FH cds 361..2740
 FT /tag= a
 FT /product=MAR subtype m1

FT misc_signal
 FT 282
 FT /tag= b
 FT /label=splice acceptor site
 FT /note=defines 5' end of exon
 FT 2780
 FT /tag= c

FT polyA_site
 FT 2780
 FT /tag= c

XX USN7241971-N.

XX 14-MAR-1989.

XX PD 08-SEP-1988; 88US-0241971.

XX PF 08-SEP-1988; 88US-0241971.

XX PR 08-SEP-1988; 88US-0241971.

XX PA (USSH) NAT INST OF HEALTH.

XX WPI; 1989-165452/22.
 DR P-PSDB; AAR06202.

XX Cloned genes for muscarinic acetylcholine receptors -
 PT for drug screening and diagnostic use.

XX PS Disclosure; ; P; English.

XX The sequence was indexed from the best available specification copy;
 CC the unidentifiable bases were annotated as 'N'.

CC The sequence may be useful for synthesis of hybridisation probes for
 CC diagnostic use. The genes are cloned by screening a rat cerebral cortex

CC CDNA library with a probe based on nucleotides 170-225 of the non-coding
 CC strand of porcine brain (m1) cDNA (modified at positions 5, 38 and 53);

CC identifying cDNA clones by hybridisation with BamHI or XhoI digests of
 CC plasmid DNA from 12 cultures contg. up to 500000 independent clones;

CC recluturing until a pool contg. less than 5000 indep. clones with a
 CC single hybridisation band are identified, and isolating individual

CC clones by colony hybridisation.

CC stable cell lines produced by transfecting Chinese hamster ovary cells
 CC (CHO-K1) with various pCD vectors contg. the gene inserts.

CC See also AAQ02068-Q020671 and AAQ00086.

CC (Note: Revised entry submitted to correct the patent number format of
 CC US Government-owned NTIS applications to prevent clashes with ongoing US

CC granted patent numbers. For further information please visit the Derwent
 CC web site at www.derwent.com/dwpl/updates/ntis.us.html.)

XX Sequence 2880 BP; 618 A; 921 C; 763 G; 550 T; 28 other;

Query Match 91.1%; Score 16.4; DB 10; Length 2880;
 Best Local Similarity 94.4%; Pred. No. 2.7e+02;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 CTCAGAGAGAGAGAGCC 18
 ||||| |||||
 Db 1230 CTCAGAGAGAGAGAGCC 1247

RESULT 15

ABK92262
 ID ABK92262 standard; DNA; 3150 BP.

AC ABK92262;
 XX

DT 15-AUG-2002 (first entry)
 XX

DE Prostate cancer-associated DNA sequence #148.
 XX

KM Prostate cancer; prostate tumour tissue; human; mammal; cytostatic;
 KW gene therapy; gene; ds.
 XX

OS Mammalia.
 OS

PN W0200230268-A2.
 PN

PD 18-APR-2002.
 PD

PF 12-OCT-2001; 2001WO-US32045.
 PF

PR 13-OCT-2000; 2000US-0687576.
 PR

PR 08-DEC-2000; 2000US-0733288.
 PR

PR 08-DEC-2000; 2000US-0733742.
 PR

PR 24-JAN-2001; 2001US-263957P.
 PR

PR 16-MAR-2001; 2001US-276791P.
 PR

PR 16-MAR-2001; 2001US-276888P.
 PR

PR 06-APR-2001; 2001US-281922P.
 PR

PR 24-APR-2001; 2001US-286214P.
 PR

PR 30-APR-2001; 2001US-0847046.
 PR

PR 04-MAY-2001; 2001US-288589P.
 PR

XX
 XX

PA (EOSB-) EOS BIO TECHNOLOGY INC.
 PA

PI Gish KC, Mack DH, Wilson KE, Afar D, Hevezi P;
 PI

XX
 XX

DR WPI; 2002-471335/50.
 DR

DR P-PSDB; ABG61943.
 DR

XX
 XX

PT Detecting a prostate cancer-associated transcript in a cell in a
 PT

PT patient, useful for diagnosing prostate cancer (PC) or screening
 PT

PT modulators of PC, by determining if prostate cancer-associated genes
 PT

XX
 XX

PS Claim 22; Page 425-426; 436pp; English.
 PS

XX
 XX

SQ Sequence 3150 BP; 653 A; 956 C; 887 G; 650 T; 4 other;
 Query Match 91.1%; Score 16.4; DB 24; Length 3150;
 Best Local Similarity 94.4%; Pred. No. 2.7e+02;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 CTCAGAGAGAGAGAGCC 18
 ||||| |||||
 Db 381 CTCAGAGAGAGAGAGCC 398

Search completed: November 9, 2002, 10:29:12
 Job time : 8.66041 secs

The present invention relates to methods of detecting a prostate cancer-associated transcript in a cell from a patient. The method comprises contacting a biological sample from the patient with prostate cancer-associated polynucleotides (designated PC genes) that selectively hybridise to a sequence that is at least 80% identical to them. The prostate cancer-associated polynucleotide sequences are differentially expressed in prostate tumour tissue or in prostate cancer and are derived from the tissues of various organisms such as humans or other mammals (e.g. mice, sheep and dogs). The methods of the invention are useful for diagnosing and treating prostate cancer in mammals. The prostate cancer-associated genes are useful for diagnosing or treating prostate cancer, as well as for identifying modulators of prostate cancer or agents that inhibit prostate cancer. The nucleic acid sequences are particularly useful in gene therapy, as a vaccine or in antisense applications.
 ABK92115-ABK92263 represent prostate cancer-associated polynucleotide sequences.

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 9, 2002, 09:10:11 ; Search time 49.0362 Seconds

(without alignments)
10682.939 Million cell updates/sec

Title: US-09-902-772-1_COPY_645_662

Perfect score: 18

Sequence: 1 ctcagagagagagagc 18

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 100%

Listing first 45 summaries

Database :
1: gb_da:*
2: gb_htg:*
3: gb_in:*
4: gb_cm:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18	100.0	1413	5	AY065661
2	18	100.0	1447	6	AY065661
3	18	100.0	170204	2	AC018539
4	18	100.0	189886	9	AC080023
5	17	94.4	71291	2	AC118263
6	17	94.4	90821	2	AC130907
7	17	94.4	121679	2	AC120952
8	17	94.4	139854	2	AC097081
9	17	94.4	147016	2	AC108185
10	17	94.4	155816	2	AC102702
11	17	94.4	165269	2	AC117064
12	17	94.4	171111	2	AC099452
13	17	94.4	181739	2	AC116059
14	17	94.4	192095	2	AP002781
15	17	94.4	195434	9	AC090744
16	17	94.4	209384	2	AC127313
17	17	94.4	229687	2	AC125404
18	17	94.4	239598	2	AC073733
19	17	94.4	295746	2	AC112789
20	17	94.4	306117	2	AC127339
21	16.4	91.1	1383	9	AF498915
22	16.4	91.1	1386	9	HSNACHR
23	16.4	91.1	1422	6	AX280889
24	16.4	91.1	1426	9	AF385587
25	16.4	91.1	1522	9	HSNML
26	16.4	91.1	1956	6	I49723
27	16.4	91.1	2100	9	HSNIMAR
28	16.4	91.1	2574	9	BC007740
29	16.4	91.1	2863	9	BC022984
30	16.4	91.1	3579	8	CBU48963
31	16.4	91.1	4132	9	HUMFNRA5
32	16.4	91.1	4204	6	AX083652
33	16.4	91.1	4204	6	HSFNRA
34	16.4	91.1	4251	6	AX281697
35	16.4	91.1	4252	9	BC008786
36	16.4	91.1	13852	10	AB026490
37	16.4	91.1	38230	8	AC067939
38	16.4	91.1	48128	6	AX059518
39	16.4	91.1	55712	2	AC100995
40	16.4	91.1	63803	2	AC120154
41	16.4	91.1	68443	2	AC124292
42	16.4	91.1	75443	2	AC118651
43	16.4	91.1	78520	9	AC004455
44	16.4	91.1	88628	2	AC099142
45	16.4	91.1	89811	2	AC108282

ALIGNMENTS

RESULT 1
LOCUS AY065661 1413 bp mRNA linear VRT 17-DEC-2001
DEFINITION Gallus gallus erg isoform C-1-1 mRNA, complete cds; alternatively spliced.
ACCESSION AY065661
VERSION AY065661.1 GI:17687440
KEYWORDS
SOURCE Gallus gallus.
ORGANISM Gallus gallus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 1413)
Iwamoto,M., Higuchi,Y., Koyama,E., Enomoto-Iwamoto,M., Kurisu,K.,

TITLE Yeh, H., Abrams, W.R., Rosenbloom, J. and Pacifici, M.
Transcription factor ERG variants and functional diversification of
Chondrocytes during limb bone development
J. Cell Biol. 150 (1), 27-40 (2000)

JOURNAL MEDLINE 20351415
PUBMED 10893254

REFERENCE 2 (bases 1 to 1413)
AUTHORS Iwamoto, M., Higuchi, Y., Enomoto-Iwamoto, M., Kurisu, K., Koyama, E.,
Yeh, H., Rosenbloom, J. and Pacifici, M.
TITLE The role of ERG (ets related gene) in cartilage development
JOURNAL Osteoarthritis. Cartil. 9 Suppl A, S41-S47 (2001)
MEDLINE 21535378
PUBMED 11680687

REFERENCE 3 (bases 1 to 1413)
AUTHORS Iwamoto, M., Higuchi, Y., Yeh, H. and Pacifici, M.
TITLE Direct Submission
JOURNAL Submitted (05-DEC-2001) Oral Anatomy and Developmental Biology,
Osaka University Faculty of Dentistry, 1-8 Yamadaoka, Suita, Osaka
565-0871, Japan

FEATURES
source Location/Qualifiers
1. .1413
/organism="Gallus gallus"
/db_xref="taxon:9031"
37. .1392
/note="alternatively spliced: transcription factor"
/codon_start=1
/product="erg isoform C-1-1"
/protein_id="AAL4089.1"
/db_xref="gi:17887441"
/translation="MASTIKELAVSSEDSLFECAVGSPLHAKTEMTASSSEYQGT
SKMSRVRQODMLSQPARVITKMECNQVNSPDSCVAKGKMSDDVNG
NYSYMEKHPPPMATNERVYIPADPTLMSSTDHYKQLEMAVKEGLPDVILF
QNIDKELCKMKTKDQFORLTPSYNADILLSHLYREKGAFFIFPNTSVYPATQRLT
TRPDLPYQARSASMTSHSHPOSATQPSSTVATKTDQRLDPYQILGPTSSRLA
NPGSQIQLMQFLLELSDSSNSNCITWEGTGEKMTDPDEVARMGESKPMYNY
DKLSRALRYDYDKNITKYGKRYKEDFGIAQLDPHPRESSMYKATPLPMSS
YHAHQKNFVAPHPALPYTSSSFFAPNPYWNPTGTYPTNLPAAHMSHLGTY
Y"

BASE COUNT 424 a 370 c 311 g 308 t
ORIGIN

Query Match 100.0%; Score 18; DB 5; Length 1413;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CTCAGAGAGAGAGAGACC 18
|||||
Db 619 CTCAGAGAGAGAGAGACC 636

RESULT 2
E31253 1447 bp DNA linear PAT 18-JUN-2001
LOCUS E31253
DEFINITION Protein having cell calcifying inhibitory activity and gene
encoding the same.
ACCESSION E31253
VERSION E31253.1 GI:13025685
KEYWORDS JP 199075871-A/1
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 1447)
AUTHORS Hiroyasu, T., Yoshinobu, H., Marijo, P., Joel, R. and Helena, E.
TITLE Protein having cell calcifying inhibitory activity and gene
encoding the same
PATENT: JP 199075871-A 1 23-MAR-1999;
CHUGAI PHARMACEUT CO LTD, HE TRUSTEES OF THE UNIVERSITY OF
PENNSYLVANIA
OS Unidentified
PN JP 199075871-A/1
PD 23-MAR-1999
PF 29-MAY-1998 JP 1998166076
PR 18-JUN-1997 US 08/878177, 20-JUN-1997 US 60/050297 PT

HIROYASU IWAMOTO, YOSHINOBU HIGUCHI, MARIJO PASHIFIKI, PI JOEL
ROZENBURG, PI HELENA E
PC C12N15/09, A61K48/00, C07K14/465, C07K16/18, C12Q1/68//A61K38/00,
PC C12P21/02,
PC C12P21/08, C12N15/00, A61K37/02
CC Strandedness: Single;
CC Topology: Linear;
FH location/Qualifiers
FT key 1. .1447
/organism="unidentified".

FEATURES
source Location/Qualifiers
1. .1447
/organism="unidentified"
/db_xref="taxon:32644"

BASE COUNT 440 a 374 c 317 g 316 t
ORIGIN

Query Match 100.0%; Score 18; DB 6; Length 1447;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CTCAGAGAGAGAGAGACC 18
|||||
Db 645 CTCAGAGAGAGAGAGACC 662

RESULT 3
AC018539 170204 bp DNA linear HTG 27-APR-2000
LOCUS AC018539
DEFINITION Homo sapiens chromosome 11 clone RP11-32702 map 11, WORKING DRAFT
SEQUENCE, 22 unordered pieces.
ACCESSION AC018539
VERSION AC018539.4 GI:7657720
KEYWORDS HTG; HTGS; PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
REFERENCE 1 (bases 1 to 170204)
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE Homo sapiens chromosome 11, clone RP11-32702
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 170204)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Beckert, R., Beda, F.,
Boguslavsky, L., Boukhalter, B., Brown, A., Castle, A., Colangelo, M.,
Collins, S., Collymore, A., Cooke, P., DeLellano, K., Dewar, K.,
Domino, M., Doyle, M., Fenesco, J., Ferrera, P., FitzHugh, W.,
Forrest, C., Gage, D., Galagan, J., Gardyna, S., Grant, G., Hagos, B.,
Heaford, A., Horton, L., Howland, J. C., Johnson, R., Jones, C., Kann, L.,
Karatas, A., Klein, J., Landers, T., Lehoczy, J., Lieu, C., Liu, G.,
Locke, K., MacDonald, P., Marquis, N., McEwan, P., McGurk, A.,
McKernan, K., Meldrum, J., Morrow, D., Naylor, J., Norman, C. H.,
O'Connor, T., O'Donnell, P., Peterson, K., Plevre, N., Pollara, V.,
Riley, R., Rothman, D., Roy, A., Santos, R., Severy, P.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Testaye, S., Theodore, J., Tirrell, A., Vassiliev, H., Viel, R., Vo, A.,
Wu, X., Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (13-DEC-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 27, 2000 this sequence version replaced gi:7143451.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1998-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L5073

```

----- Summary Statistics -----
Center clone name: 327_O_2
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 151612 bases at least Q40
Consensus quality: 159581 bases at least Q30
Consensus quality: 164253 bases at least Q20
Insert size: 161000; agarose-ftp
Insert size: 168104; sum-of-contigs
Quality coverage: 4.3 in Q20 bases; agarose-ftp
Quality coverage: 4.1 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 22 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1      1246: contig of 1246 bp in length
*      1247 1346: gap of 100 bp
*      1347 2424: contig of 1078 bp in length
*      2425 2524: gap of 100 bp
*      2525 3567: contig of 1043 bp in length
*      3568 3667: gap of 100 bp
*      3668 5470: contig of 1803 bp in length
*      5471 5570: gap of 100 bp
*      5571 7155: contig of 1585 bp in length
*      7156 7255: gap of 100 bp
*      7256 9506: contig of 2251 bp in length
*      9507 9606: gap of 100 bp
*      9607 11993: contig of 2387 bp in length
*      11994 12093: gap of 100 bp
*      12094 14691: contig of 2598 bp in length
*      14692 14791: gap of 100 bp
*      14792 17204: contig of 2413 bp in length
*      17205 17304: gap of 100 bp
*      17305 19138: contig of 1834 bp in length
*      19139 19238: gap of 100 bp
*      19239 21610: contig of 2372 bp in length
*      21611 21710: gap of 100 bp
*      21711 26767: contig of 5057 bp in length
*      26768 26867: gap of 100 bp
*      26868 34107: contig of 7240 bp in length
*      34108 34207: gap of 100 bp
*      34208 42405: contig of 8198 bp in length
*      42406 42505: gap of 100 bp
*      42506 51044: contig of 8539 bp in length
*      51045 51144: gap of 100 bp
*      51145 62544: contig of 11400 bp in length
*      62545 62644: gap of 100 bp
*      62645 76139: contig of 13495 bp in length
*      76140 76239: gap of 100 bp
*      76240 87678: contig of 11439 bp in length
*      87679 87778: gap of 100 bp
*      87779 107601: contig of 19823 bp in length
*      107602 107701: gap of 100 bp
*      107702 124522: contig of 16821 bp in length
*      124523 124622: gap of 100 bp
*      124623 143831: contig of 19209 bp in length
*      143832 143931: gap of 100 bp
*      143932 170204: contig of 26273 bp in length.

FEATURES
source
1..170204
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11"
/clone="RP11-32702"
/clone_1ib="RPCT-11 Human Male BAC"
1..1246
misc_feature

```

```

/misc_feature /note="assembly_fragment"
1347..2424 /note="assembly_fragment"
/misc_feature /note="assembly_fragment"
2525..3567 /note="assembly_fragment"
3668..5470 /note="assembly_fragment"
5571..7155 /note="assembly_fragment"
7256..9506 /note="assembly_fragment"
9607..11993 /note="assembly_fragment"
12094..14691 /note="assembly_fragment"
14792..17204 /note="assembly_fragment"
17305..19138 /note="assembly_fragment"
/misc_feature /note="assembly_fragment"
19239..21610 /note="assembly_fragment"
21711..26767 /note="assembly_fragment"
26868..34107 /note="assembly_fragment"
34208..42405 /note="assembly_fragment"
42506..51044 /note="assembly_fragment"
51145..62544 /note="assembly_fragment"
62645..76139 /note="assembly_fragment"
76240..87678 /note="assembly_fragment"
87779..107601 /note="assembly_fragment"
107702..124522 /note="assembly_fragment"
124623..143831 /note="assembly_fragment"
/misc_feature /note="assembly_fragment"
143932..170204 /note="assembly_fragment"
misc_feature /note="assembly_fragment"
47829 a 36375 c 36306 g 47565 t 2129 others
ORIGIN
Query Match 100.0%; Score 18; DB 2; Length 170204;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CTCAGAGAGAGAGAGCC 18
Db 58919 CTCAGAGAGAGAGAGCC 58936

RESULT 4
AC080023 189886 bp DNA linear PRI 23-MAY-2002
LOCUS Homo sapiens chromosome 11, clone RP11-351124, complete sequence.
DEFINITION AC080023
ACCESSION AC080023
VERSION AC080023.6 GI:21104962
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 189886)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 11, clone RP11-351124

```

JOURNAL
REFERENCE
AUTHORS

Unpublished
2 (bases 1 to 189886)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Beda,F., Boguslavsky,L.,
Bouhagalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
Choepell,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
Deatellano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P.,
Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M.,
Graham,L., Grand-Pierre,N., Hagos,B., Healdorf,A., Horton,L.,
Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., LaRoque,K.,
Lamazzes,R., Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G.,
Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McKernan,K.,
McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mienga,V.,
Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T.,
O'Donnell,P., O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K.,
Pierre,N., Pisanl,C., Pollara,V., Raymond,C., Rieback,M., Riley,R.,
Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P.,
Souarez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Straus,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
Zimmer,A. and Zody,M.

TITLE
JOURNAL
REFERENCE
AUTHORS

Submitted (23-SEP-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 189886)
Birren,B., Linton,L., Nusbaum,C., Lander,E., All,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,
Bouhagalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
Chazaro,B., Choepell,Y., Colangelo,M., Collins,S., Collymore,A.,
Cooke,A., Cooke,P., Deatellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gird,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., LaRoque,K., Lamazzes,R.,
Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G.,
Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
Nordhu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Straus,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE
JOURNAL
REFERENCE
AUTHORS

Submitted (27-APR-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 189886)
Birren,B., Linton,L., Nusbaum,C., Lander,E., All,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,
Bouhagalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
Chazaro,B., Choepell,Y., Colangelo,M., Collins,S., Collymore,A.,
Cooke,A., Cooke,P., Deatellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzgerald,M., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Gird,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I.,
Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRoque,K.,
Lamazzes,R., Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K.,
Liu,G., Maclean,C., Macdonald,P., Major,J., Marquis,N.,
Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J.,
Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C.,
Nicol,R., Nordhu,C., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C.,
Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S.,
Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Straus,N., Subramanian,A., Talamas,J., Testaye,S.,
Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,

TITLE
JOURNAL
COMMENT

Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (23-MAY-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 23, 2002 this sequence version replaced gi:20330891.
ALL repeats were identified using RepeatMasker:
Smit,A.F.A. & Green,P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L10564
Center clone name: 351_1-24

FEATURES
source

Location/Qualifiers
1..189886
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11"
/clone="Rp11-351124"
/clone_11b="RP11-11 Human Male BAC"
complement(9..251)
/rpt_family="L2"
273..875
/rpt_family="HML1"
460..571
/note="single clone coverage"
728..733
/note="<30 qual SNGL region"
738..743
/note="<30 qual SNGL region"
complement(1344..1529)
/rpt_family="MIR3"
1652..1813
/rpt_family="MIR"
3125..3212
/rpt_family="MER58B"
3242..3434
/rpt_family="AluSg/x"
3437..3694
/rpt_family="MER58B"
complement(4426..4758)
/rpt_family="L1ME2"
5217..5329
/rpt_family="L1P83"
5469..5554
/rpt_family="AT-rich"
complement(6594..6901)
/rpt_family="AluSg"
complement(6923..7108)
/rpt_family="MER2"
complement(7109..8326)
/rpt_family="r199ser3b"
complement(8327..8476)
/rpt_family="MER2"
complement(9124..9506)
/rpt_family="MIR1K"
complement(10137..10251)
/rpt_family="MIR"
10822..10932
/rpt_family="L2"
10956..11016
/rpt_family="L2"
complement(11255..11333)
/rpt_family="MIR"
11721..12146
/rpt_family="L1MA9"
12147..12218
/rpt_family="(TA)n"

```

repeat_region      12222..12638
                    /rpt_family="L1MAG"
repeat_region      12695..12741
                    /rpt_family="CATAn"
repeat_region      12876..13077
                    /rpt_family="MER21B"
unsure             complement(12953..12958)
unsure             /note="<30 qual SINGL region"
unsure             complement(12981..12985)
unsure             /note="<30 qual SINGL region"
unsure             complement(12991..12997)
unsure             /note="<30 qual SINGL region"
unsure             complement(13016..13024)
unsure             /note="<30 qual SINGL region"
unsure             complement(13053..13057)
unsure             /note="<30 qual SINGL region"
unsure             complement(13093..13099)
unsure             /note="<30 qual SINGL region"
unsure             complement(13103..13109)
unsure             /note="<30 qual SINGL region"
unsure             complement(13142..13148)
unsure             /note="<30 qual SINGL region"
unsure             complement(13836..13840)
unsure             /note="<30 qual SINGL region"
unsure             14395..14690
unsure             /note="single clone coverage"
unsure             14445..14454
repeat_region      14561..14937
                    /rpt_family="THELC"
repeat_region      15345..16167
                    /rpt_family="L2"
repeat_region      16480..16813
                    /rpt_family="MLT2C1"
repeat_region      16814..16863
                    /rpt_family="CT-rich"
repeat_region      16922..16966
                    /rpt_family="MLT2C1"
repeat_region      complement(17333..17881)
                    /rpt_family="L2"
repeat_region      18764..18787
                    /rpt_family="TTTTTA)n"

```

Query Match 100.0%; Score 18; DB 9; Length 189886;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCAGAGAGAGAGGCC 18
 DB 15122 CTCAGAGAGAGAGGCC 15139

RESULT 5
 AC118263 71291 bp DNA linear HTG 14-APR-2002
 LOCUS Mus musculus clone RP24-502N12, LOW-PASS SEQUENCE SAMPLING.
 AC118263
 VERSION AC118263.1 GI:20148143
 KEYWORDS HTG; HTGS_PHASED.
 SOURCE Mus musculus.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 71291)
 Birren, B., Linton, L., Nusbaum, C. and Lander, E.
 Mus musculus, clone RP24-502N12
 Unpublished
 2 (bases 1 to 71291)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., All, A., Allen, N.,
 Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L.,
 Boukhalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,
 Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A.,
 Cook, A., Cooke, P., Deatellano, K., Dewar, K., Diaz, J.S., Dodge, S.,

TITLE JOURNAL COMMENT

Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
 Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
 Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
 Kamat, A., Karatas, A., Kells, C., Lacroque, R., Lamazares, R.,
 Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G.,
 Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,
 McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Meus, L.,
 Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R.,
 Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,
 Oliver, J., Peterson, K., Phukhang, P., Pierre, N., Pollara, V.,
 Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
 Roman, J., Rosetti, M., Roy, A., Santos, R., Schaner, S., Schnupack, R.,
 Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
 Straus, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
 Topham, K., Travers, M., Travis, N., Triggillo, J., Vassiliev, H.,
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W., Young, G.,
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
 Submitted (14-APR-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu
 Project Information
 Center project name: L24889
 Center clone name: 502_N_12

NOTE: This record contains 88 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

1 723: contig of 723 bp in length
 * 724 823: gap of 100 bp
 * 824 1550: contig of 727 bp in length
 * 1551 1650: gap of 100 bp
 * 1651 2302: contig of 652 bp in length
 * 2303 2402: gap of 100 bp
 * 2403 3105: contig of 703 bp in length
 * 3106 3205: gap of 100 bp
 * 3206 3906: contig of 701 bp in length
 * 3907 4006: gap of 100 bp
 * 4007 4712: contig of 706 bp in length
 * 4713 4812: gap of 100 bp
 * 4813 5510: contig of 698 bp in length
 * 5511 5610: gap of 100 bp
 * 5611 6307: contig of 697 bp in length
 * 6308 6407: gap of 100 bp
 * 6408 7109: contig of 702 bp in length
 * 7110 7209: gap of 100 bp
 * 7210 7932: contig of 723 bp in length
 * 7933 8032: gap of 100 bp
 * 8033 8718: contig of 686 bp in length
 * 8719 8818: gap of 100 bp
 * 8819 9544: contig of 726 bp in length
 * 9545 9644: gap of 100 bp
 * 9645 10378: contig of 734 bp in length
 * 10379 10478: gap of 100 bp
 * 10479 11171: contig of 693 bp in length
 * 11172 11271: gap of 100 bp
 * 11272 11979: contig of 708 bp in length
 * 11980 12079: gap of 100 bp

12080 12784: contig of 705 bp in length
12785 12884: gap of 100 bp
12885 13606: contig of 722 bp in length
13607 13706: gap of 100 bp
13707 14400: contig of 694 bp in length
14401 14500: gap of 100 bp
14501 15228: contig of 728 bp in length
15229 15328: gap of 100 bp
15329 16028: contig of 700 bp in length
16029 16128: gap of 100 bp
16129 16826: contig of 698 bp in length
16827 16926: gap of 100 bp
16927 17657: contig of 731 bp in length
17658 17757: gap of 100 bp
17758 18463: contig of 706 bp in length
18464 18563: gap of 100 bp
18564 19293: contig of 730 bp in length
19294 19393: gap of 100 bp
19394 20088: contig of 695 bp in length
20089 20188: gap of 100 bp
20189 20896: contig of 708 bp in length
20897 20996: gap of 100 bp
20997 21700: contig of 704 bp in length
21701 21800: gap of 100 bp
21801 22505: contig of 705 bp in length
22506 22605: gap of 100 bp
22606 23327: contig of 722 bp in length
23328 23427: gap of 100 bp
23428 24143: contig of 716 bp in length
24144 24243: gap of 100 bp
24244 24960: contig of 717 bp in length
24961 25060: gap of 100 bp
25061 25762: contig of 702 bp in length
25763 25862: gap of 100 bp
25863 26558: contig of 696 bp in length
26559 26658: gap of 100 bp
26659 27376: contig of 718 bp in length
27377 27476: gap of 100 bp
27477 28206: contig of 730 bp in length
28207 28306: gap of 100 bp
28307 29013: contig of 707 bp in length
29014 29113: gap of 100 bp
29114 29812: contig of 699 bp in length
29813 29912: gap of 100 bp
29913 30619: contig of 707 bp in length
30620 30719: gap of 100 bp
30720 31422: contig of 703 bp in length
31423 31522: gap of 100 bp
31523 32246: contig of 724 bp in length
32247 32346: gap of 100 bp
32347 33075: contig of 729 bp in length
33076 33175: gap of 100 bp
33176 33900: contig of 725 bp in length
33901 34000: gap of 100 bp
34001 34725: contig of 725 bp in length
34726 34825: gap of 100 bp
34826 35552: contig of 727 bp in length
35553 35652: gap of 100 bp
35653 36358: contig of 706 bp in length
36359 36458: gap of 100 bp
36459 37186: contig of 728 bp in length
37187 37286: gap of 100 bp
37287 38018: contig of 732 bp in length
38019 38118: gap of 100 bp
38119 38832: contig of 714 bp in length
38833 38932: gap of 100 bp
38934 39628: contig of 696 bp in length
39629 39728: gap of 100 bp
39729 40430: contig of 702 bp in length
40431 40530: gap of 100 bp
40531 41238: contig of 708 bp in length
41239 41338: gap of 100 bp
41339 42041: contig of 703 bp in length

42042 42141: gap of 100 bp
42142 42861: contig of 720 bp in length
42862 42961: gap of 100 bp
42962 43676: contig of 715 bp in length
43677 43776: gap of 100 bp
43777 44497: contig of 721 bp in length
44498 44597: gap of 100 bp
44598 45328: contig of 731 bp in length
45329 45428: gap of 100 bp
45429 46131: contig of 703 bp in length
46132 46231: gap of 100 bp
46232 46941: contig of 710 bp in length
46942 47041: gap of 100 bp
47042 47773: contig of 732 bp in length
47774 47873: gap of 100 bp
47874 48556: contig of 683 bp in length
48557 48656: gap of 100 bp
48657 49363: contig of 707 bp in length
49364 49463: gap of 100 bp
49464 50136: contig of 673 bp in length
50137 50236: gap of 100 bp
50237 50955: contig of 719 bp in length
50956 51055: gap of 100 bp
51056 51777: contig of 722 bp in length
51778 51877: gap of 100 bp
51878 52597: contig of 720 bp in length
52598 52697: gap of 100 bp
52698 53423: contig of 726 bp in length
53424 53523: gap of 100 bp
53524 54237: contig of 714 bp in length
54238 54337: gap of 100 bp
54338 55034: contig of 697 bp in length
55035 55134: gap of 100 bp
55135 55856: contig of 722 bp in length
55857 55956: gap of 100 bp

Query Match 94.4%, Score 17, DB 2: Length 71291;
Best Local Similarity 100.0%, Pred. No. 77;
Matches 17: Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 CTCGAGAGAGAGAGC 17
Db 58992 CTCGAGAGAGAGAGC 59008
|||||

RESULT 6
AC130907 90821 bp DNA linear HNG 15-AUG-2002
LOCUS
DEFINITION
Rattus norvegicus clone CH230-268M4, *** SEQUENCING IN PROGRESS
AC130907
AC130907.1 GI:22261948
VERSION
KEYWORDS
HTG, HTGS, PHASE1.
SOURCE
Norway rat.
ORGANISM
Rattus norvegicus

REFERENCE
AUTHORS
1 (bases 1 to 90821)
Muzny D, Marie, Metzger M, Lee, Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Angilano, D.,
Anyaledechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Cesari, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durkin, K., Duval, B., Evans, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,

Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
 Gatregeorgis, E., Geer, K., Gill, R., Grady, W., Guerra, M., Guevara, W.,
 Gunatirup, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
 Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
 Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogue, M.,
 Hollins, B., Howell, S., Huliy, S., Hume, J., Idlebird, D., Jackson, A.,
 Jackson, U., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
 Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
 Kowals, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
 Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
 Lorenshewa, L., Louisedge, H., Lozado, R.J., Lu, X., Ma, J.,
 Maheshwari, M., Mahindratne, M., Mahmoud, M., Malloy, K., Mangum, A.,
 Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
 Mawhney, S., McLeod, M., McNeill, T., Meenen, E., Milosavljevic, A.,
 Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K.,
 Morris, S., Mundasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D.,
 Newton, N., Nguyen, N., Norris, S., Nwokeneme, O., Okunolu, G.,
 Olariunpagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H.,
 Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Polindexter, A.,
 Popovic, D., Primus, E., Pu, L.-L., Puzo, M., Quito, J., Rachlin, E.,
 Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y.,
 Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A.,
 Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S.,
 Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A.,
 Sisson, I., Sitter, C.D., Smaj, D., Sneed, A., Sodergren, E.,
 Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A.,
 Svatek, A., Tabors, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S.,
 Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, Y., Villasana, D.,
 Waldron, L., Walker, B., Wang, J., Wang, O., Wang, S., Warren, J.,
 Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczek, R.,
 Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S.,
 Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X.,
 Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R.,
 Holt, R.A., Smith, H.O., Weinstein, G. and Gibbs, R.A.

Unpublished
 Direct Submission
 2 (bases 1 to 90821)
 Rat Genome Sequencing Consortium.
 Direct Submission
 Submitted (15-AUG-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA

----- Genome Center -----
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information -----
 Center project name: KAOP
 Center clone name: CH230-268M4
 ----- Summary Statistics -----
 Sequencing vector: Plasmid;
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 42767 bases at least Q40
 Consensus quality: 46297 bases at least Q30
 Consensus quality: 48467 bases at least Q20

----- NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 NOTE: This is a 'working draft' sequence. It currently
 consists of 48 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.

1 1014: contig of 1014 bp in length
 * 1015 1114: gap of unknown length
 * 1115 2674: contig of 1560 bp in length
 * 2675 2774: gap of unknown length
 * 2775 4461: contig of 1687 bp in length

4462 4561: gap of unknown length
 * 4562 5578: contig of 1017 bp in length
 * 5579 5678: gap of unknown length
 * 5679 6915: contig of 1237 bp in length
 * 6916 7015: gap of unknown length
 * 7016 8872: contig of 1857 bp in length
 * 8873 8972: gap of unknown length
 * 8973 10363: contig of 1391 bp in length
 * 10364 10463: gap of unknown length
 * 10464 11732: contig of 1269 bp in length
 * 11733 11832: gap of unknown length
 * 11833 12898: contig of 1067 bp in length
 * 12899 12999: gap of unknown length
 * 12900 13000: contig of 1410 bp in length
 * 13001 14409: gap of unknown length
 * 14410 14509: gap of unknown length
 * 14510 16333: contig of 1824 bp in length
 * 16334 16434: gap of unknown length
 * 16435 17901: contig of 1468 bp in length
 * 17902 18001: gap of unknown length
 * 18002 19678: contig of 1677 bp in length
 * 19679 21150: gap of unknown length
 * 21151 21250: contig of 1372 bp in length
 * 21251 22588: gap of unknown length
 * 22589 22689: contig of 1339 bp in length
 * 22690 24585: gap of unknown length
 * 24586 24685: contig of 1896 bp in length
 * 24686 26103: gap of unknown length
 * 26104 26203: contig of 1418 bp in length
 * 26204 27328: gap of unknown length
 * 27329 27429: contig of 1125 bp in length
 * 27430 28600: gap of unknown length
 * 28601 28700: contig of 1172 bp in length
 * 28701 30492: gap of unknown length
 * 30493 30592: contig of 1792 bp in length
 * 30593 31903: gap of unknown length
 * 31904 32003: contig of 1311 bp in length
 * 32004 33786: gap of unknown length
 * 33787 33886: contig of 1783 bp in length
 * 33887 35541: gap of unknown length
 * 35542 35641: contig of 1655 bp in length
 * 35642 38037: gap of unknown length
 * 38038 38137: contig of 2396 bp in length
 * 38138 40233: gap of unknown length
 * 40234 40333: contig of 2096 bp in length
 * 40334 41990: gap of unknown length
 * 41991 42090: contig of 1657 bp in length
 * 42091 43335: gap of unknown length
 * 43336 43437: contig of 1246 bp in length
 * 43438 45178: gap of unknown length
 * 45179 45278: contig of 1742 bp in length
 * 45279 47413: gap of unknown length
 * 47414 47513: contig of 2135 bp in length
 * 47514 48603: gap of unknown length
 * 48604 48709: contig of 1096 bp in length
 * 48710 50430: gap of unknown length
 * 50431 50530: contig of 1721 bp in length
 * 50531 52901: gap of unknown length
 * 52902 53001: contig of 2371 bp in length
 * 53002 55228: gap of unknown length
 * 55229 55328: contig of 2227 bp in length
 * 55329 56869: gap of unknown length
 * 56870 58825: contig of 1541 bp in length
 * 58826 58925: gap of unknown length
 * 58926 60390: contig of 1856 bp in length
 * 60391 60490: gap of unknown length
 * 60491 63276: contig of 1465 bp in length
 * 63277 63376: gap of unknown length
 * 63377 65051: contig of 2786 bp in length
 * 65052 65151: gap of unknown length
 * 65152 67974: contig of 1675 bp in length
 * 67975 68074: contig of 2823 bp in length
 * 68075 68074: gap of unknown length

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

```

* 68075 69788: contig of 1714 bp in length
* 69789 69888: gap of unknown length
* 69889 72095: contig of 2207 bp in length
* 72095 72195: gap of unknown length
* 72195 74139: contig of 1944 bp in length
* 74139 74239: gap of unknown length
* 74239 74240: contig of 1720 bp in length
* 74240 75959: gap of unknown length
* 75959 76059: gap of unknown length
* 76059 76060: contig of 3625 bp in length
* 76060 79784: gap of unknown length
* 79784 81385: contig of 1601 bp in length
* 81385 81485: gap of unknown length
* 81485 83058: contig of 1573 bp in length
* 83058 83159: gap of unknown length
* 83159 87307: contig of 4149 bp in length
* 87307 87308: gap of unknown length
* 87308 90821: contig of 3414 bp in length.
* 87408 90821: contig of 3414 bp in length.

FEATURES
  source 1. 90821
          /organism="Rattus norvegicus"
          /db_xref="taxon:10116"
          /clone="CH230-268M4"

BASE COUNT 25931 a 16778 c 16778 g 25307 t 6027 others

ORIGIN
Query Match 94.4%; Score 17; DB 2; Length 90821;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CTCGAGAGAGAGAGAGC 17
    |||
Db 89235 CTCGAGAGAGAGAGC 89251

RESULT 7
AC120952/c
LOCUS AC120952 121679 bp DNA linear HTG 23-JUL-2002
DEFINITION Rattus norvegicus clone CH230-489C21, *** SEQUENCING IN PROGRESS
ACCESSION AC120952
VERSION AC120952.3 GI:21902734
KEYWORDS HTG; HTGS; PHASE1.
SOURCE Rattus norvegicus.
ORGANISM Rattus norvegicus.
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
          Rattus.
          1 (bases 1 to 121679)
Munzy D.M., Adams C., Adio-Oduola B., Ali-osman, F.R., Allen C.,
Albrooks S.L., Amaralunga H.C., Ate J.R., Ayele M., Banks T.,
Barbata J., Benton J., Blinag K., Blankenburg K., Bonin D.,
Bouck J., Bowle S., Brileva M., Brown E., Brown M., Bryant N.P.,
Bunay C., Burch P., Burkett C., Burrell K.L., Byrd N.C.,
Carroll T.F., Carter M., Cavazos S.R., Chacko J., Chavez D.,
Chen G., Chen R., Chen Z., Chowdhry I., Christopoulos C.,
Cleveland C.D., Cox C., Coyle M.D., Dathorne S.R., David R.,
Dayella M.L., Davis C., Davy-Carroll L., Dederich D.A.,
DeLaney K.R., Delgado O., Denn A.L., Ding Y., Dinh H.C.,
Douthwaite K.J., Draper H., Dugan-Rocha S., Durbin K.J.,
Earhart C., Edgar D., Edwards C.C., Elhaj C., Escotto M.,
Falls T., Ferraguto D., Flagg N., Ford J., Foster P., Frantz P.,
Gabisi A., Gao J., Garcia A., Garner T., Garza N., Gill R.,
Gorell J.H., Guevara W., Gunaratne P., Hale S., Hamilton K.,
Harris C., Harris K., Hart M., Havlak P., Hawes A., Hernandez J.,
Hernandez O., Hodgson A., Hogues M., Holloway C., Hollins B.,
Homsi F., Howard S., Huber J., Huliyk S., Hummel J., Jackson L.E.,
Jacobson B., Jia Y., Johnson R., Jolivet S., Joudah S.,
Karlsson E., Kelly S., Khan U., King L., Kovach J., Kovar C.,
Kretovic J., Kureshi A., Landry N., Leal B., Lewis L.C., Lewis L.,
Li J., Li Z., Lichtarge O., Lien C., Liu J., Liu W., Louisedge H.,
Lozano R.J., Lu X., Lucier A., Lucier R., Luna R., Ma J.,
Maheshwari M., Mapua P., Martin R., Martindale A., Martinez E.,
Massey E., Mawhney E., McLeod M.P., Meador M., Mel G., Metzker M.,

```

```

TITLE
JOURNAL
AUTHORS
REFERENCE
Mortley K.C.
Submitted (14-MAY-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 121679)
Mortley K.C.
Direct Submission
Submitted (23-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 18, 2002 this sequence version replaced gi:21039577.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GMRZ
Center clone name: CH230-489C21
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 79824 bases at least Q40
Consensus quality: 87081 bases at least Q30
Consensus quality: 91417 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a "working draft" sequence. It currently
* consists of 62 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1
* 1544: contig of 1544 bp in length
* 1545 1644: gap of unknown length
* 1645 2849: contig of 1205 bp in length
* 2850 2949: gap of unknown length
* 2950 4259: contig of 1310 bp in length
* 4260 4359: gap of unknown length
* 4360 5533: contig of 1174 bp in length
* 5534 5633: gap of unknown length
* 5634 6759: contig of 1126 bp in length
* 6760 6859: gap of unknown length
* 6860 7918: contig of 1059 bp in length
* 7919 8018: gap of unknown length
* 8019 9563: contig of 1545 bp in length
* 9564 9663: gap of unknown length
* 9664 10736: contig of 1073 bp in length
* 10737 10836: gap of unknown length
* 10837 12112: contig of 1276 bp in length

```

```

* 12113 12212: gap of unknown length
* 12213 13392: contig of 1180 bp in length
* 13393 13492: gap of unknown length
* 13493 14564: contig of 1072 bp in length
* 14565 14664: gap of unknown length
* 14665 15837: contig of 1173 bp in length
* 15838 15937: gap of unknown length
* 15938 18014: contig of 2077 bp in length
* 18015 18114: gap of unknown length
* 18115 19211: contig of 1097 bp in length
* 19212 19311: gap of unknown length
* 19312 20917: contig of 1606 bp in length
* 20918 21017: gap of unknown length
* 21018 22354: contig of 1337 bp in length
* 22355 22454: gap of unknown length
* 22455 23523: contig of 1069 bp in length
* 23524 23623: gap of unknown length
* 23624 24954: contig of 1331 bp in length
* 24955 25054: gap of unknown length
* 25055 26630: contig of 1576 bp in length
* 26631 26730: gap of unknown length
* 26731 27985: contig of 1255 bp in length
* 27986 28085: gap of unknown length
* 28086 29510: contig of 1425 bp in length
* 29511 29610: gap of unknown length
* 29611 30670: contig of 1060 bp in length
* 30671 30770: gap of unknown length
* 30771 32118: contig of 1348 bp in length
* 32119 32218: gap of unknown length
* 32219 33918: contig of 1700 bp in length
* 33919 34018: gap of unknown length
* 34019 35054: contig of 1036 bp in length
* 35055 35154: gap of unknown length
* 35155 36575: contig of 1421 bp in length
* 36576 36675: gap of unknown length
* 36676 38325: contig of 1650 bp in length
* 38326 38425: gap of unknown length
* 38426 39810: contig of 1385 bp in length
* 39811 39910: gap of unknown length
* 39911 41326: contig of 1416 bp in length
* 41327 41426: gap of unknown length
* 41427 42710: contig of 1284 bp in length
* 42711 42810: gap of unknown length
* 42811 44550: contig of 1740 bp in length
* 44551 44650: gap of unknown length
* 44651 46343: contig of 1693 bp in length
* 46344 46443: gap of unknown length
* 46444 48204: contig of 1761 bp in length
* 48205 48304: gap of unknown length
* 48305 49936: contig of 1632 bp in length
* 49937 50036: gap of unknown length
* 50037 52209: contig of 2173 bp in length
* 52210 52309: gap of unknown length
* 52310 53951: contig of 1642 bp in length
* 53952 54051: gap of unknown length
* 54052 55472: contig of 1421 bp in length
* 55473 55572: gap of unknown length
* 55573 57675: contig of 2103 bp in length
* 57676 57775: gap of unknown length
* 57776 59401: contig of 1626 bp in length
* 59402 59501: gap of unknown length
* 59502 60700: contig of 1199 bp in length
* 60701 60800: gap of unknown length
* 60801 63376: contig of 2576 bp in length
* 63377 63476: gap of unknown length
* 63477 65234: contig of 1758 bp in length
* 65235 65334: gap of unknown length
* 65335 67051: contig of 1717 bp in length
* 67052 67151: gap of unknown length
* 67152 69414: contig of 2263 bp in length
* 69415 69514: gap of unknown length
* 69515 71548: contig of 2034 bp in length
* 71549 71648: gap of unknown length

```

```

Query Match          94.4%; Score 17; DB 2; Length 121679;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 CTCAGAGAGAGAGGC 17
Db      94894 CTCAGAGAGAGAGGC 94878

RESULT 8
AC097081
LOCUS      AC097081
DEFINITION Rattus norvegicus clone CH230-49613, *** SEQUENCING IN PROGRESS
ACCESSION  AC097081
VERSION    AC097081.4 GI:21728501
KEYWORDS   HTG: HTGS PHASE1.
SOURCE     Norway rat.
ORGANISM   Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
            1 (bases 1 to 139854)
REFERENCE  1 (bases 1 to 139854)
AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaralunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbarella,J., Benton,J., Bimarge,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Hollway,C., Hollins,B.,
            Homsl,F., Howard,S., Huber,J., Huliy,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Licharge,O., Lien,C., Liu,J., Liu,W., Loulsedg,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mel,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohababat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokwenwo,S., Oguh,M., Okunolu,G.,
            Orangunye,N., Owiedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., RojudoKan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshitari,N., Sisson,I.,

```

Sodergren, E., Sonaika, T., Sparks, A., Stanley, H., Stone, H.,
Sutton, A., Svatek, A., Taber, P., Tamerisa, A., Tamerisa, K., Tang, H.,
Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q.,
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wallington, S.,
Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K.,
Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G., and Gibbs, R.

TITLE Direct Submission
JOURNAL 2 (bases 1 to 139854)
REFERENCE Unpublished
AUTHORS Worley, K.C.
TITLE Direct Submission
JOURNAL Submitted (09-OCT-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 139854)
Worley, K.C.
Direct Submission
Submitted (12-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 11, 2002 this sequence version replaced gi:17973481.

COMMENT ----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GHXN
Center clone name: CH230-49G13
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 9153 bases at least Q40
Consensus quality: 96388 bases at least Q30
Consensus quality: 101027 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 60 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1197: contig of 1197 bp in length
1198 1297: gap of unknown length
1298 2560: contig of 1263 bp in length
2561 2660: gap of unknown length
2661 3997: contig of 1337 bp in length
3998 4097: gap of unknown length
4098 5550: contig of 1453 bp in length
5551 5650: gap of unknown length
5651 7414: contig of 1764 bp in length
7415 7515: gap of unknown length
7515 9383: contig of 1869 bp in length
9384 9483: gap of unknown length
9484 11216: contig of 1733 bp in length
11217 11316: gap of unknown length
11317 12933: contig of 1617 bp in length
12934 13033: gap of unknown length
13034 14078: contig of 1045 bp in length
14079 14178: gap of unknown length
14179 15408: contig of 1230 bp in length
15409 15508: gap of unknown length
15509 16979: contig of 1471 bp in length
16980 17079: gap of unknown length
17080 18868: contig of 1789 bp in length
18869 18968: gap of unknown length

18969 20388: contig of 1420 bp in length
20389 20488: gap of unknown length
20489 21584: contig of 1096 bp in length
21585 21684: gap of unknown length
21685 23765: contig of 2081 bp in length
23766 23865: gap of unknown length
23866 25111: contig of 1246 bp in length
25112 25211: gap of unknown length
25212 26881: contig of 1670 bp in length
26882 26982: gap of unknown length
26983 28184: contig of 1203 bp in length
28185 28284: gap of unknown length
28285 29981: contig of 1697 bp in length
29982 30081: gap of unknown length
30082 31936: contig of 1855 bp in length
31937 32036: gap of unknown length
32037 33155: contig of 1119 bp in length
33156 33255: gap of unknown length
33256 34381: contig of 1126 bp in length
34382 34481: gap of unknown length
34482 35901: contig of 1420 bp in length
35902 36001: gap of unknown length
36002 37134: contig of 1133 bp in length
37135 37234: gap of unknown length
37235 38525: contig of 1291 bp in length
38526 38625: gap of unknown length
38626 41139: contig of 2514 bp in length
41140 41239: gap of unknown length
41240 42576: contig of 1337 bp in length
42577 42676: gap of unknown length
42677 44082: contig of 1406 bp in length
44083 44182: gap of unknown length
44183 45681: contig of 1499 bp in length
45682 45781: gap of unknown length
45782 47067: contig of 1285 bp in length
47067 47166: gap of unknown length
47167 48955: contig of 1789 bp in length
48956 49055: gap of unknown length
49055 50923: contig of 1868 bp in length
50924 51023: gap of unknown length
51024 54431: contig of 3408 bp in length
54432 54531: gap of unknown length
54532 57840: contig of 3309 bp in length
57841 57940: gap of unknown length
57941 60258: contig of 2318 bp in length
60259 60358: gap of unknown length
60359 61548: contig of 1190 bp in length
61549 61648: gap of unknown length
61649 63856: contig of 2208 bp in length
63857 63956: gap of unknown length
63957 65395: contig of 1439 bp in length
65396 65495: gap of unknown length
65496 66975: contig of 1480 bp in length
66976 67075: gap of unknown length
67076 70169: contig of 3094 bp in length
70170 70269: gap of unknown length
70270 72569: contig of 2300 bp in length
72570 72669: gap of unknown length
72670 74610: contig of 1941 bp in length
74611 74710: gap of unknown length
74711 76759: contig of 2049 bp in length
76760 76859: gap of unknown length
76860 79252: contig of 2393 bp in length
79253 79352: gap of unknown length
79353 81541: contig of 2189 bp in length
81542 81641: gap of unknown length
81642 84280: contig of 2649 bp in length
84291 84390: gap of unknown length
84391 87337: contig of 2947 bp in length
87338 87437: gap of unknown length
87438 90023: contig of 2586 bp in length
90024 90123: gap of unknown length
90124 94613: contig of 4450 bp in length

```

*      94614      94713: gap of unknown length
*      94714      97941: contig of 3228 bp in length
*      97942      98041: gap of unknown length
*      98042      100976: contig of 2935 bp in length
*      100977      101076: gap of unknown length
*      101077      104551: contig of 3475 bp in length
*      104552      104651: gap of unknown length
*      104652      106825: contig of 2174 bp in length
*      106826      106925: gap of unknown length
*      106926      109513: contig of 2588 bp in length

Query Match      94.4%; Score 17; DB 2; Length 139854;
Best Local Similarity 100.0%; Pred. No. 74;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY      2      TCACAGAGAGAGAGACC 18
          |||||
Db 102848 TCACAGAGAGAGAGACC 102864

RESULT 9
AC108185      147016 bp      DNA      linear      HTG 26-JAN-2002
LOCUS      Fells catus clone RP86-340111, WORKING DRAFT SEQUENCE, 5 unordered
DEFINITION      pieces.
AC108185
VERSION      AC108185.1 GI:18376884
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Fells catus.
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Fells.

```

```

REFERENCE
AUTHORS      Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
              Blakeley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,
              Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,
              Ho, S.-L., Idol, J.R., Karlins, E., Latic, P., Lee-Lin, S.-Q.,
              Legaspi, R., Maduro, Q.L., Maduro, V.B., Masello, C., Maskeri, B.,
              Mestrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Prasad, A.,
              Stantipop, S., Thomas, J.W., Thomas, P.J., Touchman, J.W.,
              Tsurgion, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,
              Young, A., Zhang, L. H. and Green, E.D.
              NISC Comparative Sequencing Initiative
              Unpublished
              2 (bases 1 to 147016)
              Green, E.D.
              Submitted (26-JAN-2002) NIH Intramural Sequencing Center, 8717
              Groveomont Circle, Gaithersburg, MD 20877, USA
              Center: NIH Intramural Sequencing Center
              Center code: NISC
              Web site: http://www.nisc.nih.gov
              Contact: nisc.mouse@ngl.nih.gov
              Project Information
              Center project name: coh
              Center clone name: 340111
              Summary Statistics
              Sequencing vector: plasmid; n/a; 100% of reads
              Chemistry: Dye-terminator Big Dye; 100% of reads
              Assembly program: Phrap; version 0.990319
              Consensus quality: 145202 bases at least Q40
              Consensus quality: 145456 bases at least Q30
              Consensus quality: 145735 bases at least Q20
              Insert size: 130000; agarose-efp
              Insert size: 146616; sum-of-ctctigs
              Quality coverage: 11.81x in Q20 bases; agarose-efp
              Quality coverage: 10.47x in Q20 bases; sum-of-ctctigs

```

* NOTE: This is a 'working draft' sequence. It currently consists of 5 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as

```

*      2919: contig of 2919 bp in length
*      3019: gap of unknown length
*      3020      10407: contig of 7388 bp in length
*      10408      10507: gap of unknown length
*      10508      26429: contig of 15922 bp in length
*      26430      26529: gap of unknown length
*      26530      40137: contig of 13608 bp in length
*      40138      40237: gap of unknown length
*      40238      147016: contig of 106779 bp in length.
              location/Qualifiers
              1..147016
              /organism="Fells catus"
              /db_xref="taxon:9685"
              /clone="RP86-340111"
              /clone_1fb="RP86"
              1..2919
              /note="assembly-fragment"
              3020..10407
              /note="assembly-fragment"
              10508..26429
              /note="assembly-fragment"
              26530..40137
              /note="assembly-fragment"
              clone_end:r7
              vector_side:left"
              40238..147016
              /note="assembly-fragment"
              clone_end:SP6
              vector_side:left"

```

```

BASE COUNT      45556 a 28955 c 27949 g 44136 t      420 others
ORIGIN

```

```

Query Match      94.4%; Score 17; DB 2; Length 147016;
Best Local Similarity 100.0%; Pred. No. 74;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY      1      CTCACAGAGAGAGAGC 17
          |||||
Db 109883 CTCACAGAGAGAGAGC 109899

```

```

RESULT 10
AC102702      155816 bp      DNA      linear      HTG 21-AUG-2002
LOCUS      Mus musculus clone RP24-380G7, WORKING DRAFT SEQUENCE, 24 unordered
DEFINITION      pieces.
AC102702
VERSION      AC102702.2 GI:22381692
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FUULLTOP.
SOURCE      house mouse.
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 155816)
Mus musculus, clone RP24-380G7
Unpublished
2 (bases 1 to 155816)
Britten, B., Linton, L., Nusbaum, C., Lander, E., All, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Boguslavskiy, L., Boukhalter, B.,
Brown, A., Canarata, J., Campopiano, A., Chang, J., Chazaro, B.,
Choepe, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,
Cooke, P., Dearrellano, K., Dewar, K., Diaz, J.S., Dodge, S., Fato, S.,
Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
Glade, S., Gord, S., Goyette, M., Graham, L., Hulme, W., Iliev, I., Johnson, R.,
Hagos, B., Hearford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
Jones, C., Kamat, A., Karatas, A., Kells, C., Lacroque, K.,
Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,
Maclean, C., Macdonald, P., Major, J., Margulis, N., Matthews, C.,

```



```

misc_feature      /note="assembly-fragment"
33536..33790
/note="assembly-fragment"
39891..45837
misc_feature      /note="assembly-fragment"
45938..52189
/note="assembly-fragment"
52290..57622
misc_feature      /note="assembly-fragment"
57723..64527
/note="assembly-fragment"
64628..71926
misc_feature      /note="assembly-fragment"
72027..80807
/note="assembly-fragment"
80908..91609
/note="assembly-fragment"
91710..103209
misc_feature      /note="assembly-fragment"
103310..121316
/note="assembly-fragment"
121417..155816
misc_feature      /note="assembly-fragment"
37206 a 40762 c 39429 g 36114 t 2305 others
BASE COUNT
ORIGIN

```

```

Query Match      94.4%; Score 17; DB 2; Length 155816;
Best Local Similarity 100.0%; Pred. No. 73;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 CTCGAGAGAGAGAGAGC 17
        |||
Db      94145 CTCGAGAGAGAGAGAGC 94161

```

```

RESULT 11
AC117064/c      165269 bp      DNA      1linear      HTG 18-JUL-2002
LOCUS      Rattus norvegicus clone CH230-248E3, *** SEQUENCING IN PROGRESS
DEFINITION
AC117064      165269 bp      DNA      1linear      HTG 18-JUL-2002
AC117064      165269 bp      DNA      1linear      HTG 18-JUL-2002
AC117064.3      GI:21746714
KEYWORDS      HTG; HTGS-PHASE1.
SOURCE      Norway rat.
ORGANISM      Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murine;
Rattus.
1 (bases 1 to 165269)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrook,S.L., Amaralunge,H.C., Are,J.R., Ayale,M., Banks,T.,
Barbieri,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D.,
Bouch,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carroll,T.F., Carter,M., Cavaros,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Devilla,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinb,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Einhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorelli,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
Homsj,F., Howard,S., Huber,U., Hulik,S., Hume,J., Jackson,J.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlsom,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.,
Li,J., Li,Z., Licharge,O., Lieu,J., Liu,J., Liu,W., Loulsegod,H.,
Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

```

```

Massey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokweto,S., Ogih,M., Okunodu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
Schere,S., Scott,G., Shen,H., Shoohtari,N., Slason,I.,
Sodergren,E., Sonake,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Syatek,A., Taber,P., Tamerisa,A., Tamerisa,K., Thomas,N.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Umanal,K., Vasquez,L., Vera,Y., Villalob,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 165269)
Worley,K.C.
Direct Submission
Submitted (06-APR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 165269)
Worley,K.C.
Direct Submission
Submitted (18-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 14, 2002 this sequence version replaced g1:20143512.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GUXC
Center clone name: CH230-248E3
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 1008 of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 111480 bases at least Q40
Consensus quality: 117613 bases at least Q30
Consensus quality: 122299 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 56 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
1031: contig of 1031 bp in length
1032 1131: gap of unknown length
1132 2207: contig of 1076 bp in length
2208 2307: gap of unknown length
2308 3534: contig of 1227 bp in length
3535 3634: gap of unknown length
3635 4649: contig of 1015 bp in length
4650 4749: gap of unknown length
4750 6025: contig of 1276 bp in length
6026 6125: gap of unknown length
6126 7255: contig of 1130 bp in length
7256 7356: gap of unknown length
7357 8470: contig of 1114 bp in length
8471 8569: gap of unknown length
8570 9683: contig of 1114 bp in length
9684 9783: gap of unknown length

```

* 9784 11324: contig of 1441 bp in length
* 11225 11324: gap of unknown length
* 11325 11328: contig of 1204 bp in length
* 12529 12628: gap of unknown length
* 14077 14076: contig of 1448 bp in length
* 14177 14176: gap of unknown length
* 15244 15244: contig of 1068 bp in length
* 15345 15344: gap of unknown length
* 16911 16910: contig of 1566 bp in length
* 17011 18285: contig of 1275 bp in length
* 18385 18385: gap of unknown length
* 18386 20148: contig of 1763 bp in length
* 20149 20248: gap of unknown length
* 20249 21596: contig of 1348 bp in length
* 21597 21696: gap of unknown length
* 21697 23450: contig of 1754 bp in length
* 23451 23550: gap of unknown length
* 23551 25808: contig of 2258 bp in length
* 25809 25908: gap of unknown length
* 25909 27489: contig of 1581 bp in length
* 27490 27589: gap of unknown length
* 27590 28914: contig of 1325 bp in length
* 28915 30978: contig of 1964 bp in length
* 30979 31078: gap of unknown length
* 31079 32654: contig of 1576 bp in length
* 32655 32754: gap of unknown length
* 32755 34447: contig of 1693 bp in length
* 34448 34547: gap of unknown length
* 34548 36451: contig of 1904 bp in length
* 36452 36551: gap of unknown length
* 36552 38481: contig of 1930 bp in length
* 38482 38581: gap of unknown length
* 38582 40265: contig of 1684 bp in length
* 40266 40365: gap of unknown length
* 40366 42371: contig of 2006 bp in length
* 42372 42471: gap of unknown length
* 42472 44877: contig of 2406 bp in length
* 44878 44977: gap of unknown length
* 44978 46128: contig of 1151 bp in length
* 46129 46228: gap of unknown length
* 46229 48793: contig of 2565 bp in length
* 48794 48893: gap of unknown length
* 48894 51197: contig of 2304 bp in length
* 51198 51297: gap of unknown length
* 51298 53892: contig of 2595 bp in length
* 53893 53992: gap of unknown length
* 53993 55906: contig of 1914 bp in length
* 55907 56006: gap of unknown length
* 56007 57540: contig of 1534 bp in length
* 57541 57640: gap of unknown length
* 57641 59855: contig of 2215 bp in length
* 59856 59955: gap of unknown length
* 59956 62018: contig of 2063 bp in length
* 62019 62118: gap of unknown length
* 62119 65559: contig of 3441 bp in length
* 65560 65659: gap of unknown length
* 65660 68496: contig of 2837 bp in length
* 68497 68596: gap of unknown length
* 68597 71179: contig of 2583 bp in length
* 71180 71799: gap of unknown length
* 71799 74250: contig of 2971 bp in length
* 74251 74350: gap of unknown length
* 74351 78316: contig of 3966 bp in length
* 78317 78416: gap of unknown length
* 78417 79867: contig of 1551 bp in length
* 79868 80067: gap of unknown length
* 80068 84321: contig of 4254 bp in length
* 84322 84421: gap of unknown length
* 84422 88884: contig of 4463 bp in length
* 88885 88984: gap of unknown length
* 88985 94389: contig of 5405 bp in length

Query Match 94.4%; Score 17; DB 2; Length 165269;
Best Local Similarity 100.0%; Pred. No. 73;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CTCGAGAGAGAGAGC 17
Db 87708 CTCGAGAGAGAGAGC 87692
RESULT 12
AC099452/c
LOCUS
DEFINITION
Rattus norvegicus clone CH230-62L13, *** SEQUENCING IN PROGRESS
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 171111)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaral-Lunge,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbata,J., Benton,J., Blincoe,K., Blomberg,K., Blomberg,D.,
Bouck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
DeLaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escoto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,D., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B.,
Honsi,F., Howard,S., Huber,J., Huijck,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudan,S.,
Karlsoun,E., Kelly,S., Khan,U., King,L., Kovach,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtenberg,O., Lieu,C., Liu,J., Liu,W., Louiseged,H.,
Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mel,G., Metzger,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okunnu,G.,
Orgunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Plickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,

REFERENCE AUTHORS TITLE JOURNAL	REFERENCE AUTHORS TITLE JOURNAL	REFERENCE AUTHORS TITLE JOURNAL	REFERENCE AUTHORS TITLE JOURNAL
Scherer, S., Scott, G., Shen, H., Shoostari, N., Sisson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansley, J., Taylor, C., Taylor, T., Tellrodt, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wlarczyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G., and Gldbs, R.	Worley, K.C. 2 (bases 1 to 171111) Direct Submission Submitted (15-NOV-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA 3 (bases 1 to 171111) Worley, K.C. Direct Submission Submitted (12-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA On Jul 11, 2002 this sequence version replaced gi:17973710.	Center: Baylor College of Medicine Center code: BCM Web site: http://www.hgsc.bcm.tmc.edu/ Contact: hgsc-help@bcm.tmc.edu ----- Project Information ----- Center project name: GTID Center clone name: CH230-62L13 ----- Summary Statistics ----- Sequencing vector: Plasmid; Chemistry: Dye-terminator Big Dye; 100% of reads Assembly program: Phrap; version 0.990329 Consensus quality: 118152 bases at least Q40 Consensus quality: 123041 bases at least Q30 Consensus quality: 127104 bases at least Q20 -----	* NOTE: Estimated insert size may differ from sequence length * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html) * NOTE: This is a working draft sequence. It currently * consists of 63 contigs. The true order of the pieces * is not known and their order in this sequence record is * arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence * as soon as it is available and the accession number will * be preserved. * 1 * 1247 1346: contig of 1246 bp in length * 1347 2830: gap of unknown length * 2831 2930: contig of 1484 bp in length * 2931 4009: gap of unknown length * 4010 4109: contig of 1079 bp in length * 4110 5177: gap of unknown length * 5177 5276: contig of 1067 bp in length * 5276 6506: gap of unknown length * 6506 6606: contig of 1230 bp in length * 6606 8136: gap of unknown length * 8136 8237: contig of 1530 bp in length * 8237 9269: gap of unknown length * 9269 9370: contig of 1033 bp in length * 9370 10539: gap of unknown length * 10539 10639: gap of 1170 bp in length * 10639 11852: gap of unknown length * 11852 11952: contig of 1213 bp in length * 11952 13426: gap of unknown length * 13426 13526: contig of 1474 bp in length * 13526 15102: gap of unknown length * 15102 15103: contig of 1576 bp in length * 15103 15202: gap of unknown length * 15202 15203: contig of 1798 bp in length

```

* 97773 101954: contig of 4182 bp in length
* 101955 102054: gap of unknown length
* 102055 104310: contig of 2256 bp in length
* 104311 104410: gap of unknown length
* 104411 106489: contig of 2079 bp in length
* 106490 106589: gap of unknown length
* 106590 106590: contig of 3081 bp in length
* 109671 109770: gap of unknown length
* 109771 114002: contig of 4232 bp in length
* 114003 114102: gap of unknown length
* 114103 118233: contig of 4131 bp in length

Query Match      94.4% Score 17: DB 2: Length 171111;
Best Local Similarity 100.0% Fred. No. 73;
Matches 17: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCAGAGAGAGAGAGC 17
    |||||
Db 93770 CTCAGAGAGAGAGAGC 93754

RESULT 13
AC116059/c
LOCUS
DEFINITION Rattus norvegicus clone CH230-211M21, *** SEQUENCING IN PROGRESS
ACCESSION AC116059
VERSION AC116059.2 GI:21732024
KEYWORDS HTG: HTGS. PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
          Rattus.
          1 (bases 1 to 181739)
          Muzny,D.M., Adams,C., Agio-Oduola,B., Ali-osman,F.R., Allen,C.,
          Alstrooks,S.L., Amarantunge,H.C., Are,J.R., Ayele,M., Banks,T.,
          Barreira,J., Benton,J., Blamage,K., Blankenburg,K., Bonnin,D.,
          Bouck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
          Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
          Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
          Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
          Cleeland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
          Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
          Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
          Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
          Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escoto,M.,
          Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frintz,P.,
          Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
          Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
          Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
          Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
          Homi,F., Howard,S., Huber,J., Huij,S., Hume,J., Jackson,L.E.,
          Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
          Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
          Kratovic,U., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
          Li,J., Li,Z., Lichtenarge,O., Liew,C., Liu,J., Liu,W., Louiseded,H.,
          Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
          Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
          Massey,E., Mawhney,E., Mcleod,M.P., Meador,M., Mei,G., Metzger,M.,
          Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
          Moser,M., Neal,D., Newton,J., Newton,S., Oguh,M., Okunou,G.,
          Nguyen,N., Nickerson,E., Nwokenko,S., Oguh,M., Okunou,G.,
          Oranuyne,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
          Peters,L., Plickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
          Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G.,
          Scherer,S., Scott,G., Shen,H., Shooshitari,N., Sisson,T.,
          Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H.,
          Sutton,J., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
          Tatey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
          Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
          Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
          Williams,G., Williamson,A., Wlarczyk,R., Woodson,S., Worley,K.,

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 181739)
Worley,K.C.
Direct Submission
Submitted (24-MAR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 181739)
Worley,K.C.
Direct Submission
Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 11, 2002 this sequence version replaced gi:19698586.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GLKY
Center clone name: CH230-211M21
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 127648 bases at least Q40
Consensus quality: 134343 bases at least Q30
Consensus quality: 138969 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 79 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1
* 1048 1047: contig of 1047 bp in length
* 1148 1147: gap of unknown length
* 2382 2381: contig of 1234 bp in length
* 2482 2481: gap of unknown length
* 3659 3658: contig of 1177 bp in length
* 3759 3758: gap of unknown length
* 5264 5263: contig of 1505 bp in length
* 5364 5363: gap of unknown length
* 6434 6433: contig of 1071 bp in length
* 6435 6434: gap of unknown length
* 6535 6425: contig of 1881 bp in length
* 8426 8525: gap of unknown length
* 8526 8525: gap of unknown length
* 9530 9529: contig of 1004 bp in length
* 9629 9629: gap of unknown length
* 9630 11305: contig of 1676 bp in length
* 11306 11405: gap of unknown length
* 11406 12827: contig of 1422 bp in length
* 12828 12827: gap of unknown length
* 12928 12927: gap of unknown length
* 13937 13936: contig of 1009 bp in length
* 14036 14036: gap of unknown length
* 14037 15284: contig of 1248 bp in length
* 15285 15384: gap of unknown length
* 15385 16409: contig of 1025 bp in length
* 16410 16509: gap of unknown length
* 17984 17983: contig of 1474 bp in length
* 18084 19646: gap of unknown length
* 19647 19746: gap of unknown length
* 19747 20789: contig of 1563 bp in length
* 20790 20889: gap of unknown length

```

```

* 20890 23058: contig of 2169 bp in length
* 23059 23158: gap of unknown length
* 23159 24531: contig of 1373 bp in length
* 24532 24631: gap of unknown length
* 24632 26563: contig of 1932 bp in length
* 26564 26663: gap of unknown length
* 26664 29170: contig of 2507 bp in length
* 29171 29270: gap of unknown length
* 30467 30567: contig of 1197 bp in length
* 30568 31579: contig of 1012 bp in length
* 31580 31679: gap of unknown length
* 31680 32980: contig of 1301 bp in length
* 32981 33080: gap of unknown length
* 33081 34862: contig of 1782 bp in length
* 34863 34962: gap of unknown length
* 34963 36065: contig of 1103 bp in length
* 36066 36165: gap of unknown length
* 36166 37381: contig of 1216 bp in length
* 37382 37481: gap of unknown length
* 37482 39092: contig of 1611 bp in length
* 39093 39192: gap of unknown length
* 39193 40789: contig of 1597 bp in length
* 40790 40889: gap of unknown length
* 40890 42715: contig of 1826 bp in length
* 42716 42815: gap of unknown length
* 42816 44433: contig of 1618 bp in length
* 44434 44533: gap of unknown length
* 44534 46048: contig of 1515 bp in length
* 46049 46148: gap of unknown length
* 46149 47725: contig of 1577 bp in length
* 47726 47825: gap of unknown length
* 47826 49506: contig of 1681 bp in length
* 49507 49606: gap of unknown length
* 49607 51836: contig of 2230 bp in length
* 51837 51936: gap of unknown length
* 51937 53402: contig of 1466 bp in length
* 53403 53502: gap of unknown length
* 53503 55056: contig of 1554 bp in length
* 55057 55156: gap of unknown length
* 55157 57252: contig of 2096 bp in length
* 57253 57352: gap of unknown length
* 57353 58899: contig of 1547 bp in length
* 58900 58999: gap of unknown length
* 59000 60532: contig of 1533 bp in length
* 60533 60632: gap of unknown length
* 60633 62095: contig of 1463 bp in length
* 62096 62195: gap of unknown length
* 62196 63728: contig of 1533 bp in length
* 63729 63828: gap of unknown length
* 63829 65484: contig of 1656 bp in length
* 65485 65584: gap of unknown length
* 65585 67315: contig of 1731 bp in length
* 67316 67415: gap of unknown length
* 67416 68918: contig of 1503 bp in length
* 68919 69018: gap of unknown length
* 69019 71011: contig of 1993 bp in length
* 71012 71111: gap of unknown length
* 71112 73135: contig of 2024 bp in length
* 73136 73235: gap of unknown length
* 73236 75430: contig of 2195 bp in length
* 75431 75530: gap of unknown length
* 75531 77778: contig of 2248 bp in length
* 77779 77878: gap of unknown length
* 77879 79632: contig of 1754 bp in length
* 79633 79732: gap of unknown length
* 79733 81175: contig of 1443 bp in length
* 81176 81275: gap of unknown length
* 81276 83305: contig of 2030 bp in length
* 83306 83405: gap of unknown length
* 83406 85611: contig of 2206 bp in length
* 85612 85711: gap of unknown length
* 87075: contig of 1364 bp in length

```

```

* 87076 87175: gap of unknown length
* 87176 88201: contig of 1026 bp in length
* 88202 88301: gap of unknown length
* 88302 90394: contig of 2093 bp in length

Query Match          94.48; Score 17; DB 2; Length 181739;
Best Local Similarly 100.0%; Pred. No. 73;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Caps 0;

OY      2 TCAGAGAGAGGAGCC 18
Db 171767 TCAGAGAGAGGAGCC 171751

```

```

RESULT 14
AP002781
LOCUS
DEFINITION
Homo sapiens chromosome 11 clone RP11-660K15 map 11q, WORKING DRAFT
SEQUENCE, 16 unordered pieces.
ACCESSION
AP002781
VERSION
AP002781.1 GI:9188619
KEYWORDS
HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE
Homo sapiens DNA, clone:RP11-660K15.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

```

```

REFERENCE
1 (bases 1 to 192095)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Homo sapiens 192,095 genomic DNA of 11q
Published Only in Database (2000)
2 (bases 1 to 192095)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
Submitted (10-JUL-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555, Japan
(E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
tel:81-42-778-9923, Fax:81-42-778-9924)

```

COMMENT

```

----- Genome Center
Center: RIKEN Genomic Sciences Center(GSC)
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@gsc.riken.go.jp
----- Project Information
Center project name: Humdrat11
Center clone name: RP11-660K15
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 180110 bases at least Q40
Consensus quality: 186278 bases at least Q30
Consensus quality: 189098 bases at least Q20
Insert size: 190595; sum-of-contigs
Quality coverage: 5.13x in Q20 bases; sum-of-contigs
-----
NOTE: This is a 'working draft' sequence. It currently consists of
16 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs N, but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved
1
44030 43929 contig of 43929 bp in length
72010 71909 contig of 27880 bp in length
92966 92865 contig of 20856 bp in length
113294 113193 contig of 20228 bp in length
129044 128843 contig of 15650 bp in length
144281 144281 contig of 15238 bp in length
144382 153828 contig of 9447 bp in length
153929 162906 contig of 8978 bp in length

```

	ch	94.4%	Score 17:	DB 2:	Length 192095;	
	1 Similarity	100.0%	Pred. No. 72;			
	17; Conservative	0;	Mismatches	0;	Indels	0; Gaps
CAGAGAGAGAGGAGACC	18					
CACAGAGAGAGGAGGCC	20684					
AC090744	195434 bp	DNA	linear	PRI 01-DEC-2001		
Homo sapiens chromosome 11, clone RP11-660K15, complete sequence.						
AC090744	GI:17223210					
HTG.						
Homo sapiens.						
Homo sapiens						
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;						
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.						
1 (bases 1 to 195434)						
Birren,B., Linton,L., Nusbaum,C., Allen,N., Anderson,S.,						
Barna,N., Bastien,V., Boguslavsky,L., Bouckgalter,B., Brown,A.,						
Camarat,J., Campopiano,A., Chepel,Y., Colangelo,M., Collins,S.,						
Collymore,A., Cooke,P., Deakellano,K., Dewar,K., Diaz,J.S.,						
Dodge,S., Fero,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J.,						
Gardyna,S., Glnde,S., Goette,M., Graham,L., Grand-Pierre,N.,						
Hagds,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,						
Jones,C., Karatas,A., LaRoque,K., Lamazares,R., Landers,T.,						
Lehoczky,J., Levine,R., Liu,G., Maclean,C., MacDonald,P.,						
Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,						
McPheters,R., Meidrim,J., Meneus,L., Mihova,T., Mienga,V.,						
Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,						
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,						
Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,						
Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Roselli,M.,						
Roy,A., Santos,R., Schauer,S., Schuback,R., Seaman,S., Severy,P.,						
Sounez,C., Spencer,R., Stange-Thomann,N., Stojanovic,N.,						
Struss,M., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,						
Trawers,N., Travis,N., Trigillo,J., Vassiliev,H., Viel,R., Vo,A.,						
Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,						
Zembek,L., Zimmer,A. and Zody,M.						
Submitted (09-MAR-2001) Whitehead Institute/MIT Center for Genome						
Research, 320 Charles Street, Cambridge, MA 02141, USA						
3 (bases 1 to 195434)						
Birren,B., Linton,L., Nusbaum,C., Lander,E., All,A., Allen,N.,						
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Bouckgalter,B.,						
Brown,A., Camarat,J., Campopiano,A., Chang,J., Chiaro,B.,						
Chapel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,						
Cooke,P., Deakellano,K., Dewar,K., Diaz,J.S., Dodge,S., Fero,S.,						
Ferreita,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,						
Gerdes,A., Gord,S., Goette,M., Graham,L., Grand-Pierre,N.,						
Hagds,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,						
Jones,C., Karatas,A., LaRoque,K., Lamazares,R., Landers,T.,						
Lehoczky,J., Levine,R., Liu,G., Maclean,C., MacDonald,P.,						
Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,						
McPheters,R., Meidrim,J., Meneus,L., Mihova,T., Mienga,V.,						
Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,						
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,						
Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,						
Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Roselli,M.,						
Roy,A., Santos,R., Schauer,S., Schuback,R., Seaman,S., Severy,P.,						
Sounez,C., Spencer,R., Stange-Thomann,N., Stojanovic,N.,						
Struss,M., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,						
Trawers,N., Travis,N., Trigillo,J., Vassiliev,H., Viel,R., Vo,A.,						
Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,						
Zembek,L., Zimmer,A. and Zody,M.						

repeat_region /rpt_family="AluJb"
32942 .33075 /rpt_family="(TA)n"
repeat_region complement(33076 .33183)
/rpt_family="AluJb"
repeat_region 33184 .33224
/rpt_family="MSRA"
repeat_region complement(33358 .33554)

Query Match 94.4%; Score 17; DB 9; Length 195434;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TCAGAGGAGGAGGCC 18
|||||
Db 122772 TCAGAGGAGGAGGCC 122788

Search completed: November 9, 2002, 11:41:29
Job time : 382.036 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 9, 2002, 09:11:56 ; Search time 3067.84 Seconds

(without alignments)
7638.889 Million cell updates/sec

Title: US-09-902-772-1

Perfect score: 1447

Sequence: 1 gaattccgcgaacgaataa.....gaaagaagcgcgaagaaa 1447

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

EST: *
1: em_estba: *
2: em_esthum: *
3: em_estlin: *
4: em_estlun: *
5: em_estlov: *
6: em_estlpl: *
7: em_estro: *
8: em_hnc: *
9: gb_estl: *
10: gb_est2: *
11: gb_hnc: *
12: gb_est3: *
13: gb_est4: *
14: gb_est5: *
15: em_estlun: *
16: em_estlom: *
17: gb_gss: *
18: em_gss_hum: *
19: em_gss_inv: *
20: em_gss_pln: *
21: em_gss_vrt: *
22: em_gss_fun: *
23: em_gss_mam: *
24: em_gss_mus: *
25: em_gss_other: *
26: em_gss_pro: *
27: em_gss_rtd: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	485.4	33.5	880	14	BQ233264 AGENCOURT
2	482.2	33.3	715	12	BG388025 602412867
3	475.4	32.9	642	13	BT13036 1601408.Y
4	468.6	32.4	701	9	AU136709
5	468.6	32.4	1089	14	BQ212127 AGENCOURT
6	462	31.9	634	13	BM489636 pgmzn.pk0

7	452.2	31.3	658	10	BB660034	BB660034
8	445.2	30.8	915	12 <th>BG390291</th> <th>BG390291</th>	BG390291	BG390291
9	444.2	30.7	928	12 <th>BG259459</th> <th>BG259459</th>	BG259459	BG259459
10	442	30.5	1045	13 <th>BM456833</th> <th>BM456833</th>	BM456833	BM456833
11	436.2	30.1	675	10 <th>BB639043</th> <th>BB639043</th>	BB639043	BB639043
12	432.6	29.9	629	10 <th>BB577887</th> <th>BB577887</th>	BB577887	BB577887
13	428.8	29.6	629	10 <th>AW948986</th> <th>AW948986</th>	AW948986	AW948986
14	425.6	29.4	522	13 <th>BM088063</th> <th>BM088063</th>	BM088063	BM088063
15	380	26.3	827	9 <th>AJ456498</th> <th>AJ456498</th>	AJ456498	AJ456498
16	367.8	25.4	522	12 <th>BF566705</th> <th>BF566705</th>	BF566705	BF566705
17	367.2	25.4	946	13 <th>BG691566</th> <th>BG691566</th>	BG691566	BG691566
18	352.8	24.4	946	13 <th>BI558012</th> <th>BI558012</th>	BI558012	BI558012
19	335	23.2	553	12 <th>BG256864</th> <th>BG256864</th>	BG256864	BG256864
20	317.2	21.9	671	13 <th>BI558530</th> <th>BI558530</th>	BI558530	BI558530
21	313.4	21.7	740	9 <th>AJ456506</th> <th>AJ456506</th>	AJ456506	AJ456506
22	313	21.6	1074	14 <th>BQ953977</th> <th>BQ953977</th>	BQ953977	BQ953977
23	308	21.3	666	12 <th>BG390572</th> <th>BG390572</th>	BG390572	BG390572
24	304.8	21.1	645	10 <th>BB656030</th> <th>BB656030</th>	BB656030	BB656030
25	298.2	20.6	651	10 <th>BB442803</th> <th>BB442803</th>	BB442803	BB442803
26	297.8	20.6	1245	14 <th>BQ212133</th> <th>BQ212133</th>	BQ212133	BQ212133
27	297.4	20.6	524	9 <th>AJ1794504</th> <th>AJ1794504</th>	AJ1794504	AJ1794504
28	293	20.2	420	13 <th>BI181694</th> <th>BI181694</th>	BI181694	BI181694
29	284.6	19.7	674	10 <th>BB238163</th> <th>BB238163</th>	BB238163	BB238163
30	261.6	18.1	1015	9 <th>AL554108</th> <th>AL554108</th>	AL554108	AL554108
31	259.8	18.0	400	10 <th>BE245360</th> <th>BE245360</th>	BE245360	BE245360
32	256.2	17.7	619	10 <th>AW421310</th> <th>AW421310</th>	AW421310	AW421310
33	241	16.7	512	12 <th>BE831331</th> <th>BE831331</th>	BE831331	BE831331
34	232.6	16.1	327	10 <th>AW209658</th> <th>AW209658</th>	AW209658	AW209658
35	232.4	16.1	634	10 <th>BB655125</th> <th>BB655125</th>	BB655125	BB655125
36	228.6	15.8	465	14 <th>R87572</th> <th>R87572</th>	R87572	R87572
37	218.8	15.1	551	9 <th>AA467723</th> <th>AA467723</th>	AA467723	AA467723
38	215.8	14.9	815	9 <th>AJ446143</th> <th>AJ446143</th>	AJ446143	AJ446143
39	214.2	14.8	1068	14 <th>BQ062565</th> <th>BQ062565</th>	BQ062565	BQ062565
40	211	14.6	520	10 <th>BE233460</th> <th>BE233460</th>	BE233460	BE233460
41	209	14.4	756	9 <th>AJ445352</th> <th>AJ445352</th>	AJ445352	AJ445352
42	208	14.4	872	14 <th>BQ232274</th> <th>BQ232274</th>	BQ232274	BQ232274
43	201.4	13.9	788	9 <th>AJ456869</th> <th>AJ456869</th>	AJ456869	AJ456869
44	200.2	13.8	675	13 <th>BM440515</th> <th>BM440515</th>	BM440515	BM440515
45	198	13.7	874	14 <th>BQ215477</th> <th>BQ215477</th>	BQ215477	BQ215477

ALIGNMENTS

RESULT 1
BQ233264
LOCUS
DEFINITION
AGENCOURT_7565836 NIH_MGC_92 Homo sapiens cdna clone IMAGE:6041526
5', mRNA sequence.
ACCESSION
BQ233264
VERSION
BQ233264.1 GI:20414664
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
NIH-MGC <http://mgs.nci.nih.gov/>.
1 (bases 1 to 880)
TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ARCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
Plate: LRAM13279 row: h column: 07
High quality sequence stop: 608.
Location/Qualifiers
I. .880

FEATURES

source

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6041526"
/clone_lib="NIH_MGC_92"
/tissue_type="embryonal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/Note="Organ: testis; Vector: pCMV-Sport6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally; oligo-dr primed. Average insert size 2.5 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."

```

```

BASE COUNT      250 a      231 c      207 g      189 t      3 others
ORIGIN

```

```

Query Match      33.5%; Score 485.4; DB 14; Length 880;
Best Local Similarity 84.4%; Pred. No. 9e-139;
Matches 546; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

```

```

QY      8 GCGAAGCAATTAATTTATTAAGCAATTAATTAAGCAATTAATTTATGACATTTATGGC 67
      19 GAGAACCAATTAATTTATTAATTAATTAATTAATTAATTAATTAATTAATTTATGAGC 78
QY      68 AAGCACTATTAAGAGCAATTAATTAATTAATTAATTAATTAATTAATTTATGAGTGC 127
      79 CAGCACTATTAAGAGCAATTAATTAATTAATTAATTAATTAATTAATTTATGAGTGC 138
QY      128 CTACGAGTATGAGCAATTAATTAATTAATTAATTAATTAATTAATTTATGAGTGC 187
      139 CTACGAGTATGAGCAATTAATTAATTAATTAATTAATTAATTAATTTATGAGTGC 198
QY      188 GGAACATCAAGATTAATTAATTAATTAATTAATTAATTAATTAATTTATGAGTGC 247
      199 ACAGACTTCAAGATTAATTAATTAATTAATTAATTAATTAATTAATTTATGAGTGC 258
QY      248 CAGAGTATTAAGATTAATTAATTAATTAATTAATTAATTAATTAATTTATGAGTGC 307
      259 CAGGTCACATTAATTAATTAATTAATTAATTAATTAATTAATTTATGAGTGC 318
QY      308 TGATGACTGAGTATTAATTAATTAATTAATTAATTAATTAATTTATGAGTGC 367
      319 TGATGACTGAGTATTAATTAATTAATTAATTAATTAATTAATTTATGAGTGC 378
QY      368 GAACATTAAGATTAATTAATTAATTAATTAATTAATTAATTTATGAGTGC 427
      379 GAACATTAAGATTAATTAATTAATTAATTAATTAATTAATTTATGAGTGC 438
QY      428 ACAGAGTATTAAGATTAATTAATTAATTAATTAATTAATTTATGAGTGC 487
      439 GCGCAGTATTAAGATTAATTAATTAATTAATTAATTTATGAGTGC 498
QY      488 GCTGAGTATTAAGATTAATTAATTAATTAATTAATTTATGAGTGC 547
      499 GCTGAGTATTAAGATTAATTAATTAATTAATTTATGAGTGC 558
QY      548 CATGATTAAGATTAATTAATTAATTAATTAATTTATGAGTGC 607
      559 CATGATTAAGATTAATTAATTAATTAATTTATGAGTGC 618
QY      608 CTATAGAGTATTAATTAATTAATTAATTTATGAGTGC 654
      619 CTATAGAGTATTAATTAATTAATTTATGAGTGC 665

```

```

RESULT 2
BG388025      715 bp      mRNA      linear      EST 12-MAR-2001
LOCUS      602412867.1 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:4521233.5,
DEFINITION      mRNA sequence.
ACCESSION      BG388025
VERSION      BG388025.1 GI:13281471
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens

```

```

REFERENCE
AUTHORS      Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
TITLE      NIH-MGC http://mgs.nci.nih.gov/.
JOURNAL      National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT      Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cga@b-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM10420 row: f column: 18
High quality sequence stop: 693.

```

```

FEATURES
source
location/Qualifiers
1..715
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4521233"
/clone_lib="NIH_MGC_92"
/tissue_type="embryonal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/Note="Organ: testis; Vector: pCMV-Sport6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally; oligo-dr primed. Average insert size 2.5 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."

```

```

BASE COUNT      206 a      188 c      172 g      149 t
ORIGIN

```

```

Query Match      33.3%; Score 482.2; DB 12; Length 715;
Best Local Similarity 84.1%; Pred. No. 8e-138;
Matches 544; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

```

```

QY      8 GCGAAGCAATTAATTTATTAAGCAATTAATTAAGCAATTAATTTATGACATTTATGGC 67
      39 GAGAACCAATTAATTTATTAATTAATTAATTAATTAATTAATTTATGAGC 98
QY      68 AAGCACTATTAAGAGCAATTAATTAATTAATTAATTAATTTATGAGTGC 127
      99 CAGCACTATTAAGAGCAATTAATTAATTAATTAATTTATGAGTGC 158
QY      128 CTACGAGTATGAGCAATTAATTAATTAATTAATTTATGAGTGC 187
      159 CTACGAGTATGAGCAATTAATTAATTAATTTATGAGTGC 218
QY      188 GGAACATCAAGATTAATTAATTAATTAATTTATGAGTGC 247
      219 ACAGACTTCAAGATTAATTAATTAATTTATGAGTGC 278
QY      248 CAGAGTATTAAGATTAATTAATTAATTTATGAGTGC 307
      279 CAGGTCACATTAATTAATTAATTTATGAGTGC 338
QY      308 TGATGACTGAGTATTAATTAATTAATTTATGAGTGC 367
      339 TGATGACTGAGTATTAATTAATTTATGAGTGC 398
QY      368 GAACATTAAGATTAATTAATTAATTTATGAGTGC 427
      399 GAACATTAAGATTAATTAATTTATGAGTGC 458
QY      428 ACAGAGTATTAAGATTAATTAATTTATGAGTGC 487
      459 GCGCAGTATTAAGATTAATTTATGAGTGC 518
QY      488 GCTGAGTATTAAGATTAATTTATGAGTGC 547
      519 GCTGAGTATTAAGATTAATTTATGAGTGC 578

```


Query Match	32.4%	Score 468.6;	DB 9;	Length 701;
Best Local Similarity	84.0%;	Pred. No. 1.3e-133;		
Matches 539;	Conservative 0;	Mismatches 102;	Indels 1;	Gaps 1;
QY 8	GGGAACGAAATATTTATTTATTTAGCAATTTATTTAGCAATCAATAATCTTGATCAGATTATGGC 67			
Db 30	GAGAAGCAATAATTTATTTATTTAAATTTATTTAAATTTAAATCAATAAATCTTGATTTGATTTGGC 89			
OY 68	AAGCACTTTAAGAAAGCATTTACAGTGTGATGATGAGAACAGCAGTCTCTGTTGAGTGTGC 127			
Db 90	CAGCACTTTAAGAAAGCCTTTATCAGTGTGATGATGAGAACAGCAGTGTGTTGATGATGTC 149			
OY 128	CTACGAGTGGCCGCCCTTGCAAGACAGAAATGACAGCCTCTCTTCAGTGAATATGG 187			
Db 150	CTACGAGACGCCACACCTGGCTTAAAGACAGATGACCGCTCTCTTCAGGAGCTATATGG 209			
OY 188	GCAAAATCAATCAAAAGATGAGACCGCGCGTCCCGACAGAGATGTTATATACAGACCCCGGC 247			
Db 210	ACAGACTTCCAAAGAGAGACCCACCGGCTCCCTCAGAGATTTGGCTGTCAACCCCGACG 269			
OY 248	CAGAGTTACCTTAATGATGAGTGTATACCCCAAAACAGGTTAATGGGTCAAGGAATTCAC 307			
Db 270	CAGGCTCACCATCAAAATGAGATGTATACCTTATGCCAGGTGAATGGCTCAAGGAATCTTCC 329			
OY 308	TGATGACTGCAGCCGTGGCAAAAAGAGGGAATGGTTATGACAGTTCAGCAAAATGTTGGAT 367			
Db 330	TGATGAATGACAGTGTGGCCAAAGCGGGGAATGGTTGGGACGCCAGACACCGTGGGAT 389			
OY 368	GAACATATGGAAGCTACATGGAAGAGAGACATTTTCGGCTCCAAATATGACACCAATGA 427			
Db 390	GAACATACGGCAGCTTCATGAGAGAGAGAACACATCTGCCACCCCAAAACATGACACGACGA 449			
OY 428	ACGAAGATTTATTTGTCGACAGATCTTACGTTATGAGAGACAGACCAATGTACGGCAGTG 487			
Db 450	GCGCAGATTTATCGTGCACAGATCTTACGCTATGAGATACAGACATGTGCGGCAATG 509			
OY 488	GCTGGAGTGGGCACTGAAGAGATATGCTTCCACAGCTGAGCATCTTGTGTTCCAGAA 547			
Db 510	GCTGGAGTGGGCGGTGAAGAATATATGGCTTCCACAGCTCAAAACATCTTGTATTTCCAGA 569			
OY 548	CATGATGAGGGAAGAGTGTGTAAATATACCAA-GATGACTTCCAGAGACTCAGCCGA 606			
Db 570	CATGATGAGGGAAGAACATGTGCACAAATATACCAAGAGAGACTTCNAAGANCTCAACCCCA 629			
OY 607	GCTATTAACGAGATTAATCTCTCTGTCACAGCTACACTACTCTCA 648			
Db 630	GCTACACAGCCGACATCTTCTCTACATCTTCTACACTACTCTCA 671			

BASE COUNT

199 a 185 c 166 g 145 t 6 others

ORIGIN

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="PLACE1004911"
/clone_1b="PLACE1"
/tissue_type="Placenta"
/Note="Vector: pME185FL3"

RESULT 5

BO212127

LOCUS

DEFINITION

AGNCNCURT_7571287 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:6052136

ACCESSION

BO212127

VERSION

BO212127.1

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 1089)

AUTHORS

NIH-MGC <http://mgc.ncl.nih.gov/>.

TITLE

National Institutes of Health, Mammalian Gene Collection (MGC)

1089 bp

mrna

linear

EST 02-MAY-2002

[illegible]

```

|||||
Db      621  GCTACATGGCCGACATCTCTCTCACAATCTCCACTACCTCAGAGAGA 668

RESULT 6
LOCUS    BM489636
DEFINITION
        BM489636                634 bp      mRNA      linear      EST 07-FEB-2002
        pm2n.pk011.124 Normalized Chicken Breast Muscle, Leg Muscle, and
        Epiphyseal Growth Plate cDNA library (pm2n) Gallus gallus cDNA
        clone pm2n.pk011.124 5' similar to sp|Q90837|ERG_CHICK
        TRANSCRIPTIONAL REGULATOR ERG p1r|S60754 transcription factor erg -
        chicken emb|CA54404.1| (X77159) ERG [Gallus gallus], mRNA
        sequence.
ACCESSION
        BM489636
VERSION   BM489636.1  GI:18610567
KEYWORDS  EST.
SOURCE    chicken.
ORGANISM  Gallus gallus
           Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
           Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
           Phasianinae; Gallus.
REFERENCE
        1 (bases 1 to 634)
AUTHORS   Cogburn, L.A. and Monsonego-Ornan, E.
TITLE      ESTs from Normalized Chicken Breast Muscle, Leg Muscle, and
           Epiphyseal Growth Plate cDNA library, USDA/IRAFs Animal Genome
           Project
JOURNAL   Unpublished (2002)
COMMENT    Contact: Larry A. Cogburn
           University of Delaware
           Townsend Hall, Newark, DE 19717, USA
           Tel: 302-831-1335
           Fax: 302-831-2822
           Email: cogburn@udel.edu, www.chickest.udel.edu.
FEATURES
   source
       1..634
           /organism="Gallus gallus"
           /strain="Commercial broiler and Ottawa Res. Centre
           Strains 90 & 21"
           /db_xref="taxon:9031"
           /clone="pm2n.pk011.124"
           /clone_1lb="Normalized Chicken Breast Muscle, Leg Muscle,
           and Epiphyseal Growth Plate cDNA library (pm2n)"
           /issue_type="Breast muscle, leg muscle and epiphyseal
           growth plate"
           /dev_stage="Breast, leg: Embryo(d19): post-hatch(1d,1,3,5,7,9,
           11 weeks): growth plate(1d,7d,14d post-hatch)"
           /lab_host="E. coli EMD108"
           /note="Vector: PCWVSPORT6; Library made from equivalent
           pools of total RNA isolated from each tissue (embryonic
           muscle 33.3%; juvenile muscle 33.3%; and epiphyseal growth
           plate 33.3% of the final RNA pool). Single pass sequencing
           from 5'-end"
BASE COUNT      192 a      164 c      138 g      140 t
ORIGIN
Query Match      31.9%; Score 462; DB 13; Length 634;
Best Local Similarity 87.2%; Pred. No. 1.4e-131;
Matches 553; Conservative 0; Mismatches 0; Indels 81; Gaps 1;

```

```

|||||
Db      181  TCACGCCGAGCTATTAACGACAGATATCCTCTGTACACACTTACCTCAGAGAGACTC 240
        655  -----
        241  CTTCTCACATTTTACCTTCAGATGATGTGATAAGCCCTTACAAAACCTCCACGGTTAA 300
        655  -----GAGGACCCACTTTTATTTTTCCAATATACATCACTTTACCAG 696
        301  TGCATGCTAGAAACACAGACAGAGACCACTTTATTTTCCAAATATACATCACTTTACCAG 360
        697  AAGCAAGCCAAAGAAATTAACAAACAGGCCAGATTACTTATGACACAGCGAGGAGATCAG 756
        361  AAGCAAGCCAAAGAAATTAACAAACAGGCCAGATTACTTATGACACAAACGAGGAGATCAG 420
        757  CGTGAGAGCTACAGCCATCCACATCAGTCAAAAGCTAACCCACATCACTTTCAACAG 816
        421  CGTGAGAGCTACAGCCATCCACATCAGTCAAAAGCTAACCCACATCACTTTCAACAG 880
        817  TGCCCAAAACAGAAAGACAGCGCTCCTCAGTTAGATCCTTATCAGATTTCTTGACCGACCA 876
        481  TGCCCAAAACAGAAAGACAGCGCTCCTCAGTTAGATCCTTATCAGATTTCTTGACCGACCA 540
        877  GCACCGCTCTTGCAATTCACAGGAGGTGGCGCAGATACACATTCGACGTTCTTCTGACG 936
        541  GCACCGCTCTTGCAATTCACAGGAGGTGGCGCAGATACACATTCGACGTTCTTCTGACG 600
        937  TTCTGTGGAGAGCTCCCAACTCCAACTGATCAG 970
        601  TTCTGTGGAGAGCTCCCAACTCCAACTGATCAG 634

```

```

RESULT 7
LOCUS    BB660034
DEFINITION
        BB660034 RIKEN full-length enriched, 13 days embryo lung Mus
        musculus cDNA clone D430027104 5', mRNA sequence.
ACCESSION
        BB660034
VERSION   BB660034.1  GI:16493855
KEYWORDS  EST.
SOURCE    house mouse.
ORGANISM  Mus musculus
           Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
           1 (bases 1 to 668)
REFERENCE
        Arakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanasaki, T., Hara, A.,
        Hiramoto, K., Hori, F., Ishii, Y., Ito, M., Kawai, J., Konno, H., Kouda,
        M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ohno, M.,
        Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki,
        D., Shibata, K., Shingawa, A., Shiraki, T., Sogabe, Y., Suzuki, H.,
        Tagami, M., Tagawa, A., Takehashi, F., Takeda, Y., Tanaka, T., Toya, T.,
        Muramatsu, M. and Hayashizaki, Y.
        RIKEN Mouse ESTs (Arakawa, T., et al. 2001)
TITLE      Unpublished (2001)
JOURNAL   Contact: Yoshihide Hayashizaki
COMMENT    Laboratory for Genome Exploration Research Group, RIKEN Genomic
           Sciences Center(GSC), Yokohama Institute
           The Institute of Physical and Chemical Research (RIKEN)
           1-7-22 Saitama-cho, Tsukuba, Ibaraki, 305-8575, Japan
           Tel: 81-45-503-9222
           Fax: 81-45-503-9216
           Email: genome-res@gsr.riken.go.jp/
           URL: http://genome.gsc.riken.go.jp/
           Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh,
           M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
           Normalization and subtraction of cap-trapper-selected cDNAs to
           prepare full-length cDNA libraries for rapid discovery of new
           genes. Genome Res. 10 (10), 1617-1630 (2000)
           wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,
           Matsui, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura,
           S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and
           Hayashizaki, Y.
           RIKEN integrated sequence analysis (RISA) system--384-format

```



```
QY 138 CCCACCTTGGAAGAGAGAAATGACAGCCTCTCTTCCAGTGAATATGGCCAAACATCA 197
    || || || || || || || || || || || || || || || || || || || || || ||
Db 134 CCACACCTGGGTAAAGACAGAGATGACCGCGCTCTCTCCAGCAGTATGACAGACTTCC 193
QY 198 AAGATGAGCCCGCGGCTTCCCGACGAGACGTATACAGCGCCCGCGCAGGTTACC 257
    || || || || || || || || || || || || || || || || || || || || || ||
Db 194 AAGATGAGCCCGCGGCTTCCCGACGAGATGGTGTCTCAACCCCGACGCGAGGTCACC 253
QY 258 ATTAAGATGAGAGTGAACCCAAACAGGTTATGGTCAAGGAATTCACCTGATGATGC 317
    || || || || || || || || || || || || || || || || || || || || || ||
Db 254 ATCAAAATGGAGATTAACCTCAGCAGTGAATGGCTCAAGAACTCTCTGATGAATGC 313
QY 318 AGCGTGGCAAAAGAGAGGAAATGGTTAGCATTCAGACATGTTGGGATGAATGGA 377
    || || || || || || || || || || || || || || || || || || || || || ||
Db 314 AGTGTGGCCAAAGGCGGGAAGATGGTGGGACGCCACAGACCCGTTGGGATGAATCAGGC 373
QY 378 AGCTACATGGAAGAGAGCATATTCGCCCTCCAAATATGACACCATGAACGAAGATT 437
    || || || || || || || || || || || || || || || || || || || || || ||
Db 374 AGCTACATGGAAGAGAGCATATTCGCCCTCCAAATATGACACCATGAACGAAGATT 432
QY 438 ATTTGTCACAGC-AGATCTCTAGTATGAGACAGACATGATGAGGAGTGGAGTGG 496
    || || || || || || || || || || || || || || || || || || || || || ||
Db 433 ATGCTGCGCAGCAGATCTCTAGTATGAGTACAGACCATGTGGCGAGTGGCTGGAGTG 492
QY 497 GGCAGTGAAGAGATGATGCTCTCCAGACGTGGACATGTTGTTCCAGAGACATTTGATG 556
    || || || || || || || || || || || || || || || || || || || || || ||
Db 493 GGGCGGTAAAGAAATATGCTCTCCAGACGTCAACATCTTTATTCAGAAACATCGATGG 552
QY 557 GAAAGATTTGTTGTTAAATGACCAAGATGACTTCCAGAGACTCCGCGGAGCTTAACGC 616
    || || || || || || || || || || || || || || || || || || || || || ||
Db 553 GAAGAGACTGTGGCAAGATGACCAAGAGACGACTTCCAGAGGCTCCCGCAGCTCAACATGC 612
QY 617 AGATATCTCTCTGTACACCTACACTACTCTCAGAGAGA 654
    || || || || || || || || || || || || || || || || || || || || || ||
Db 613 CGACAT-CTTCTCTCACATCTCCACTACTCTCAGAGAGA 649

RESULT 9
Bg259459 928 bp mRNA linear EST 13-FEB-2001
LOCUS 602378556F1 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:4509541 5',
DEFINITION mRNA sequence.
ACCESSION Bg259459
VERSION Bg259459.1 GI:12769275
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 928)
AUTHORS NIH-MGC http://mgc.ncl.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@rs-femail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
plate: L1AM10389 row: 0 column: 14
High quality sequence stop: 717.
Location/Qualifiers
1..928
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4509541"
/clone_lib="NIH_MGC_92"
/tissue_type="embryonal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
JOURNAL
```

```
/note="Organ: testis; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 2.5 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC library."
BASE COUNT 265 a 242 c 220 g 201 t
ORIGIN
Query Match 30.7%; Score 444.2; DB 12; Length 928;
Best Local Similarity 83.5%; Pred. No. 5.5e-126;
Matches 540; Conservative 0; Mismatches 103; Indels 4; Gaps 3;

QY 8 GCGAACAATATATTTATAGCAATTTATGACATCAATATCTTGATCAATTAATGC 67
    || || || || || || || || || || || || || || || || || || || || || ||
Db 33 GAGAAACAATATATTTATTTAACTTTATTAACATTAATTAACCTGATTCATTAATGCG 92
QY 68 AAGCATATTTAAGAACATTTATCAGTGTGAGGAAAGACCAAGTCTGTTGAGTGGC 127
    || || || || || || || || || || || || || || || || || || || || || ||
Db 93 CAGCATATTTAAGAACCTTTATCAGTGTGAGGAGACCAAGT-CTGTTTGAAGTGTGC 151
QY 128 CTACGATGCGCCCGCTTGCAAGAGACAGAAATGACAGCCTCTCTCCAGTGAATATGG 187
    || || || || || || || || || || || || || || || || || || || || || ||
Db 152 CTACGAAACGCCACACCTGCTAAGACAGATACCGCTCTCTCCAGCAGTATGG 211
QY 188 GCAAAATCAAAAGATGAGCCCGCGTTCGCCAGCAGACTGTTATTCACAGCCCGCGC 247
    || || || || || || || || || || || || || || || || || || || || || ||
Db 212 ACAGACTTCCAAAGATGAGCCCGCGTTCGCCAGCAGATGGTGTCTCAACCCCGAGC 271
QY 248 CAGAGTTACATTTAAGATGGATGTTAACCACAGGTTAATGGTCAAGAAATTCACC 307
    || || || || || || || || || || || || || || || || || || || || || ||
Db 272 CAG-GTCAACCAATCAAAATGAGATGTAACCTTAGCAGAGTGAATGCTCAAGAACTCTCC 330
QY 308 TGATGACTGCGAGCTGGCAAAAGAGGAAATGTTAGACAGTTCAGACATGTTGGGAT 367
    || || || || || || || || || || || || || || || || || || || || || ||
Db 331 TGATGAATGCAAGTGTGGCAAAAGCGGGAGATGTTGGCAGCCACACCGTTGGGAT 390
QY 368 GAAGTATGAAGCTTACATGGAAGAAAGCATATTCGCGCTCCAAATATGACACCAATGA 427
    || || || || || || || || || || || || || || || || || || || || || ||
Db 391 GAAGTACGGCAGCTACATGAGGAGAAAGCATGACACCCCAACATGACACGAAGA 450
QY 428 ACGAAGATTAATTTGCGACAGATCTCTAGCTTTGAGACACACATGACAGGAGTG 487
    || || || || || || || || || || || || || || || || || || || || || ||
Db 451 GCGCAGATTAATTCGTGCGACAGCATCTACGCTTGAAGATGACACCATGTCGCGAGTG 510
QY 488 GCTGAGTGGGAGAGAGATGATGCTTCCAGAGCTGACATCTTGTGTTCCAGAA 547
    || || || || || || || || || || || || || || || || || || || || || ||
Db 511 GCTGAGTGGGCGGTGACAGAAATATGGCTTCCAGAGCTCAACATCTTGTATTCAGAA 570
QY 548 CATTGATGGAAAGAGTGTGTAATGACCAAGATGACTTCCAGAGACTCAGCCGAG 607
    || || || || || || || || || || || || || || || || || || || || || ||
Db 571 CATCGATGGGAAGAACTGTGCAAGATGACCAAGAGAGACTTCCAGAGGCTCACCCC-A 628
QY 608 CTATAAGCAGATATCTCTCTGTACACCTACACTACTACTCAGAGAGA 654
    || || || || || || || || || || || || || || || || || || || || || ||
Db 629 GTACAATGCGGACATCTCTCTCAATCTTCCACTACTACTCAGAGAGA 675

RESULT 10
Bg456833 1045 bp mRNA linear EST 05-FEB-2002
LOCUS Bg456833
DEFINITION AGENCOURT_6404041 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:5583753
5', mRNA sequence.
ACCESSION Bg456833
VERSION Bg456833.1 GI:18505873
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 1045)
AUTHORS NIH-MGC http://mgc.ncl.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
```


QY 379 GCTACATGAGAGAGATATTCGGCTCCAAATATGACCAACCATGACGAGAGTTA 438
 Db 320 GCTACATGAGAGAGAGATATTCGGCTCCAAACATGACCAACGAGGAGAGTTA 379
 QY 439 TTGTGCCAGCAGATCCTACGTTATGAGACAGACAGACATGATGAGGAGTGGAGTGG 498
 Db 380 TCGTGCCAGCAGATCCTACGTTATGAGAGACAGACATGATGAGGAGTGGAGTGG 439
 QY 499 CAGTGAAGAGATGATGCTTTCAGACGATGACATCTTGTTCGAGAAATGATGGA 558
 Db 440 CGGTGAAGAGATGATGCTTTCAGACGATGACATCTTGTTCGAGAAATGATGGA 499
 QY 559 AAGATGTTGTAATGACCAAGATGATGCTTTCAGACGATGACGAGTATTAACGAG 618
 Db 500 AGGAACTGTGAGATGATGCTTTCAGACGATGACATCTTGTTCGAGAAATGATGGA 559
 QY 619 ATATCTCTGTGACACCTACATCTACCTACAGAGAGA 654
 Db 560 ACATCTCTGTGACATCTACATCTACATCTACAGAGAGA 595

RESULT 14
 LOCUS BM088063 592 bp mRNA linear EST 19-NOV-2001
 DEFINITION 501436 MARC 2BOV Bos taurus cDNA 5', mRNA sequence.
 ACCESSION BM088063
 VERSION BM088063.1 GI:16998691
 KEYWORDS EST.
 SOURCE cow.
 ORGANISM Bos taurus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
 Bovidae; Bovinae; Bos.
 1 (bases 1 to 592)

REFERENCE
 AUTHORS Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
 Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett,
 G.L., Heaton,M.P., Laegreid,M.W., Rohrer,G.A., Chitko-McKown,C.G.,
 Pettee,G., Holt,I., Karanycheva,S., Liang,F., Quackenbush,J. and
 Keeler,J.W.
 TITLE Sequence evaluation of four pooled-tissue normalized bovine cDNA
 libraries and construction of a gene index for cattle
 JOURNAL Genome Res. 11 (4), 626-630 (2001)
 MEDLINE 21180013
 COMMENT Contact: Smith TPL
 USDA, ARS, US Meat Animal Research Center
 PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402 762 4366
 Fax: 402 762 4390
 Email: smith@email.marc.usda.gov
 Single pass sequencing. Bases called and alt_trimmed with phred
 v0.980904.e. Vector identified by cross_match with the -minscore 18
 and -mismatch 12 options.
 PCR Primers
 FORWARD: AGGAACACGATGACCAT
 BACKWARD: GTTTCGCCAGTCACGACG
 Plate: 139 row: D column: 21
 Seq primer: ATTGAGTGACACTATG.
 location/Qualifiers
 1..592

FEATURES

source
 1..592
 /organism="Bos taurus"
 /db_xref="taxon:9913"
 /clone_id="MARC 2BOV"
 /tissue_type="pooled"
 /lab_host="DH10B"
 /note="Vector: PCMV SPORT6; Site 1: NotI; Site 2: SalI;
 library made from pooled tissue from testis, thymus,
 semitendinosus muscle, longissimus muscle, pancreas,
 adrenal, and endometrium."
 BASE COUNT 152 a 177 c 159 g 104 t
 ORIGIN

Query Match 29.4%; Score 425.6; DB 13; Length 592;
 Best Local Similarity 82.4%; Pred. NO. 2.6e-120;

Matches 488; Conservative 0; Mismatches 104; Indels 0; Gaps 0;
 QY 52 TTGATACATTATGGCAGACATTTATGAGAGCATTTATGATGAGTGAAGACAGT 111
 Db 1 TTGATGATTATGGCCAGCATTTATGAGAGCTTTATGATGAGTGAAGACAGT 60
 QY 112 CTTGTTGAGTGTGCTTACGAGATCGCCCACTTGCAGAGACAGAAATGACAGCTTCT 171
 Db 61 CACTGTTGAGATGCGCTTACGAGACAGCGCACCTCCCAAGACAGATGACCGGCTCT 120
 QY 172 CTTCAGTGAATATGGGCAACATCAAGATGAGCGCGGCTTCCAGAGAGAGTGT 231
 Db 121 CCTCGGCGACTATGAGACAGATCAAGATGAGCGCGGCTTCCAGAGAGAGTGT 180
 QY 232 TATCAGACCCCGCGCAGAGTTACCATTAAGATGAGTGTATACCAACAGGTTAATG 291
 Db 181 TGTCTCAACCCCGACGAGGATCAACATCAAGATGAGTGTATACCAACAGGTTAATG 240
 QY 292 GGTCAAGGAATTCACCTTATGATGATGACGCGTCAAAAGAGAGGAAATGTTAGCTT 351
 Db 241 GCTCAAGGAATTCACCTTATGATGATGATGATGATGATGATGATGATGATGATG 300
 QY 352 CAGACAAATGTTGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 411
 Db 301 CGGACACCGGTGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
 QY 412 ATATGACAAACCAATGAGAGAGATTTATGAGCAGATCCTACGTTATGAGACAGAG 471
 Db 361 ACATGACCAACCAATGAG 420
 QY 472 ACCATGACGAGCAGTGGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 531
 Db 421 ACCATGTCGGGAG 480
 QY 532 TCTTGTGTTTCCAGAAACATTTATGAGGAGAGAGTGTGTAAATATGACCAATGACTTCC 591
 Db 481 TCTTATTTATTCAGAAATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540
 QY 592 AGAGACTCAGCCGCGAGATTAACGAGATATCTCTCTACACCTACACTA 643
 Db 541 AGAGGCTACCCCGAGAGCTAACAAGCTGATCTTCTCTGACACTCCACTA 592

RESULT 15
 LOCUS AJ456498 827 bp mRNA linear EST 22-APR-2002
 DEFINITION AJ456498 riken1 Gallus gallus cDNA clone bml3r2, mRNA sequence.
 ACCESSION AJ456498
 VERSION AJ456498.1 GI:2026594
 KEYWORDS EST.
 SOURCE chicken.
 ORGANISM Gallus gallus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 Phasianinae; Gallus.
 1 (bases 1 to 827)

REFERENCE
 AUTHORS Buerstedde,J.M.
 TITLE Gallus gallus bursal lymphocyte EST
 JOURNAL Unpublished (2002)
 COMMENT Contact: Buerstedde JM
 Cellular Immunology
 Heinrich-Pette-Institute
 Martinistr. 52, 20251 Hamburg, Germany
 Email: URL: http://genetics.hpi.uni-hamburg.de/dt40est.html.
 location/Qualifiers
 1..827

FEATURES

source
 1..827
 /organism="Gallus gallus"
 /db_xref="taxon:9031"
 /clone="bml3r2"
 /clone_id="riken1"
 /cell_type="bursal lymphocyte"
 /dev_stage="2-3 weeks old"
 /note="CB indred strain"

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 9, 2002, 11:41:36 ; Search time 112.599 Seconds

(without alignments)
4559.304 Million cell updates/sec

Title: US-09-902-772-1

Perfect score: 1447
Sequence: 1 gaattccgcgaacgaataat.....gaaagaagcgcaagaanaa 1447Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 320260 seqs, 177392727 residues

Total number of hits satisfying chosen parameters: 640520

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications_NA:*

- 1: /cgn2_6/ptodata1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata1/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata1/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata1/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata1/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata1/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	387	26.7	567	10	US-09-864-761-18410
2	354.6	24.5	473	10	US-09-864-761-20472
3	276.6	19.1	549	10	US-09-923-779-54
4	255	17.6	420	10	US-09-864-761-3705
5	255	17.6	454	10	US-09-864-761-1653
6	180	12.4	1752	10	US-09-850-799-1
7	167.8	11.6	553	10	US-09-864-761-7427
8	163.2	11.3	348	10	US-09-864-761-29364
9	163.2	11.3	477	10	US-09-864-761-5035
10	163.2	11.3	519	10	US-09-864-761-12798
11	162.2	11.2	225	10	US-09-864-761-21761
12	122.4	8.5	165	10	US-09-864-761-23294
13	122.4	8.5	1884	10	US-09-925-300-420
14	122.2	8.4	2188	10	US-09-920-300A-1716
15	122.2	8.4	2188	12	US-10-033-528-1716
16	122.2	8.4	2268	10	US-09-920-300A-1693
17	122.2	8.4	2268	12	US-10-033-528-1693
18	115.8	8.0	3178	9	US-10-108-605-124
19	115.8	8.0	3178	9	US-10-108-605-128

20	113	7.8	2952	10	US-09-925-297-257	Sequence 257, App
21	96	6.6	2333	10	US-09-920-300A-1788	Sequence 1788, App
22	96	6.6	2333	10	US-09-880-107-3316	Sequence 3316, App
23	96	6.6	2333	12	US-10-033-528-1788	Sequence 1788, App
24	70.4	4.9	4190	10	US-09-962-832-111	Sequence 111, App
25	66.2	4.6	414	10	US-09-864-761-10426	Sequence 10426, A
26	65.8	4.5	303	10	US-09-864-761-22055	Sequence 22055, A
27	63.2	4.4	518	9	US-10-046-935-233	Sequence 233, App
28	59.8	4.1	852	10	US-09-759-143-44	Sequence 44, App1
29	59.8	4.1	852	10	US-09-780-669-44	Sequence 44, App1
30	59.8	4.1	852	10	US-09-030-606-44	Sequence 44, App1
31	59.8	4.1	852	10	US-09-822-827-44	Sequence 44, App1
32	59.8	4.1	852	10	US-09-115-453-44	Sequence 44, App1
33	59.8	4.1	1426	10	US-09-925-297-309	Sequence 309, App
34	59.8	4.1	1429	10	US-09-764-864-320	Sequence 320, App
35	59.8	4.1	5045	9	US-09-974-298-12	Sequence 12, App1
36	58.2	4.0	440	10	US-09-960-352-11873	Sequence 11873, A
37	55.2	3.8	135	10	US-09-783-590-12136	Sequence 12136, A
38	55.2	3.8	488	10	US-09-777-564-468	Sequence 468, App
39	55.2	3.8	507	10	US-09-777-564-489	Sequence 489, App
40	53.6	3.7	451	10	US-09-998-598-32	Sequence 32, App1
41	53.6	3.7	459	10	US-09-998-598-32	Sequence 2290, App
42	53.6	3.7	1915	10	US-09-964-824A-101	Sequence 101, App
43	53.6	3.7	1915	10	US-09-964-824A-563	Sequence 563, App
44	53.6	3.7	1915	10	US-09-880-107-3420	Sequence 3420, App
45	53.6	3.7	1915	10	US-09-967-768A-192	Sequence 192, App

ALIGNMENTS

RESULT 1
US-09-864-761-18410/c
Sequence 18410, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Neomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30

OY	897	GGGAGTGGGAGATACAGCGATATGGGAGTTCTCTCTGAGGACTTCTGTGGAGAGCTCCAC	956
Db	473	GGCAGTGGCCAGATCCAGCTTTTGGAGTTCTCTCTGAGACTCTGTGGAGAGCTCCAC	414
OY	957	TCCAACTGCATCACTGCTGGGAGGGCCACAAATGGGAGTTTAAAGTGCACACCCCTATGAA	1018
Db	413	TCCAGCTGCATCACTGGGAGGACCAACGGGAGATTCAAGATGACGATCCGACAG	354
OY	1017	GTGGCTGGGCGTTGGGAGAGAGAAAAGCAAACTAATGATATATGACAACTCAGC	1078
Db	353	GTGGCCGGGGCTGGGAGGCGGAAGCAAAACCAACATGAACTACATTAAGTCCAGC	294
OY	1077	CGTCACTTCGCTACTACTATGACAAAATAATATGACTAAAGTTCATGTTAAAGCTAT	1138
Db	293	CGGCGCTCGGTACTACTATGACAAAGAAATATATACCAAGTTCATGGAAAGGCTAC	234
OY	1137	GCCTACAAATTTGATTTCCACGGAAATGCTCAGGCGCTCCAGCTCCAGCTCCAGATCA	1198
Db	233	GCTACACAGTTCCGATTCACAGGGATTCGCCCAAGCCCTCCAGCCCAACCCCGGAGTCA	174
OY	1197	TCCATGTACAAATCCCATAGAGACTCCCTCATATGATTTCTACCATCAACGCCCGAG	1258
Db	173	TCTCTGTACAGTACCCCTAGACTCCCTCATATGAGGCTCTCATATAGCCCAACCCACAG	114
OY	1257	AAGATGAACTTTGTAGCTCCCCATCCCCCTGCTTTCGCCGTAACTCAATCCAGCTTTT	1318
Db	113	AAGATGAACTTTGTGGGCCCAACCCCTCAGGCCCTCCCGTGAACCTTCAGATTTTTT	54
OY	1317	GCTGCCCTAATCATCTCTGGAATTCACCACTGAGAGCATCTACCCCAATAC	1369
Db	53	GCTGCCCAAAACCATCTCTGGAATTCACCAACTGAGGGGATATATACCCCAACAC	1

```

RESULT 3
; US-09-923-779-54
; Sequence 54, Application US/09923779
; Patent No. US20020076721A1
; GENERAL INFORMATION:
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF PANCREATIC CANCER
; FILE REFERENCE: 210121.553
; CURRENT APPLICATION NUMBER: US/09/923,779
; CURRENT FILING DATE: 2001-08-06
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 54
; LENGTH: 549
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 356, 363, 417, 433, 442, 461, 463, 464, 469, 479, 485, 489,
; LOCATION: 537, 545
; OTHER INFORMATION: n = A,T,C or G
US-09-923-779-54

Query Match          19.1%; Score 276.6; DB 10; Length 549;
Best Local Similarity 81.2%; Pred. Mismatches 76; Indels 1; Gaps 1
Matches 332; Conservative 0;

QY      1011 GATGAAGTGGCGCTGGCGTGGGAGAGGAAACCAACCTAACATGAACTATGACAA 1070
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB       2 GAGCAGGTGGCCCGCGCGCGGGGAGAGCGCAAGCAACCCCAACCTGAACTACGATAAG 61

QY      1071 CTCACCCCGTGCCTTGCCTACTACTATGACAAATAATATGACTAAATTCATGGTAA 1130
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB       62 CTCACGCCGGCCCTCGCTTACTACTATGACAAAGAACATATATGACCAAGTCCATGGGAAG 121

QY      1131 CGCTATGCCCTCAAAATTGATTTCCACGGAATCGCTCAGGCCCTCCAGCCTCACCCCTCA 1190

```

Db	122	CGCTACGGCTCAAGTTTCGACTTCCAGAGGGATGCCGACGGCCCTCCAGGCCGCCACCCCGC	181
Qy	1131	GAATCATTCATGTTCAAAATACCATCAGACTTCCTCAATGATGTTCTTACCATGACAC	1250
Db	182	GAGTCACTCTGTACAACTACCCCTCAGACTTCCTCATCATGGGCTCTTATACGCCAC	241
Qy	1251	CCCCAAGATGAACTTTGTAGTCCCATCCCTCGCTTTGGCCGTAACTCATCCAGC	1310
Db	242	CCACAGAAGATGAACCTTTGGGGCCGCCACCTCCAGCCCTCCCGTGACATCTCCAGT	301
Qy	1311	TTTTTTGGTCCCTTAATTCATCTGGAATTCACCAACTGGAGGCTCAGCCCCAATAC	1370
Db	302	TTTTTTGGTCCCAAAACCATACTGGAAATTCACCAACTGGGGGATATATACCCCTTCACT	361
Qy	1371	AGGCTGCAGTGTCTATATAGCTTCCCATCT-TGGCACTCACTACTAA	1418
Db	362	ANGCTCCCAACAGCCATATAGCCCTTTTCATCTGGGGCACTTCACTCTAA	410

```

1      RESULT 4
2      US-09-864-761-3705/C
3      ; Sequence 3705, Application US/09864761
4      ; Patent No. US20020048763A1
5      ; GENERAL INFORMATION:
6      ; APPLICANT: Penn, Sharon G.
7      ; APPLICANT: Rank, David R.
8      ; APPLICANT: Hanzel, David K.
9      ; APPLICANT: Chen, Wensheng
10     ; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
11     ; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
12     ; FILE REFERENCE: Aecomica-X-1
13     ; CURRENT FILING DATE: US/09/864,761
14     ; CURRENT FILING DATE: 2001-05-23
15     ; PRIOR APPLICATION NUMBER: US 60/180,312
16     ; PRIOR FILING DATE: 2000-02-04
17     ; PRIOR APPLICATION NUMBER: US 60/207,456
18     ; PRIOR FILING DATE: 2000-05-26
19     ; PRIOR APPLICATION NUMBER: US 09/632,366
20     ; PRIOR FILING DATE: 2000-08-03
21     ; PRIOR APPLICATION NUMBER: GB 24263.6
22     ; PRIOR FILING DATE: 2000-10-04
23     ; PRIOR APPLICATION NUMBER: US 60/236,359
24     ; PRIOR FILING DATE: 2000-09-27
25     ; PRIOR APPLICATION NUMBER: PCT/US01/00666
26     ; PRIOR FILING DATE: 2001-01-30
27     ; PRIOR APPLICATION NUMBER: PCT/US01/00667
28     ; PRIOR FILING DATE: 2001-01-30
29     ; PRIOR APPLICATION NUMBER: PCT/US01/00664
30     ; PRIOR FILING DATE: 2001-01-30
31     ; PRIOR APPLICATION NUMBER: PCT/US01/00669
32     ; PRIOR FILING DATE: 2001-01-30
33     ; PRIOR APPLICATION NUMBER: PCT/US01/00665
34     ; PRIOR FILING DATE: 2001-01-30
35     ; PRIOR APPLICATION NUMBER: PCT/US01/00668
36     ; PRIOR FILING DATE: 2001-01-30
37     ; PRIOR APPLICATION NUMBER: PCT/US01/00663
38     ; PRIOR FILING DATE: 2001-01-30
39     ; PRIOR APPLICATION NUMBER: PCT/US01/00662
40     ; PRIOR FILING DATE: 2001-01-30
41     ; PRIOR APPLICATION NUMBER: PCT/US01/00661
42     ; PRIOR FILING DATE: 2001-01-30
43     ; PRIOR APPLICATION NUMBER: PCT/US01/00670
44     ; PRIOR FILING DATE: 2001-01-30
45     ; PRIOR APPLICATION NUMBER: US 60/234,667
46     ; PRIOR FILING DATE: 2000-09-21
47     ; PRIOR APPLICATION NUMBER: US 09/608,408
48     ; PRIOR FILING DATE: 2000-06-30
49     ; PRIOR APPLICATION NUMBER: US 09/774,203
50     ; PRIOR FILING DATE: 2001-01-29
51     ; NUMBER OF SEQ ID NOS: 49117
52     ; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
53     ; SEQ ID NO 3705

```

```

1      : PRIOR FILING DATE: 2001-01-30
2      : PRIOR APPLICATION NUMBER: PCT/US01/00669
3      : PRIOR FILING DATE: 2001-01-30
4      : PRIOR APPLICATION NUMBER: PCT/US01/00665
5      : PRIOR FILING DATE: 2001-01-30
6      : PRIOR APPLICATION NUMBER: PCT/US01/00665
7      : PRIOR FILING DATE: 2001-01-30
8      : PRIOR APPLICATION NUMBER: PCT/US01/00668
9      : PRIOR FILING DATE: 2001-01-30
10     : PRIOR APPLICATION NUMBER: PCT/US01/00663
11     : PRIOR FILING DATE: 2001-01-30
12     : PRIOR APPLICATION NUMBER: PCT/US01/00662
13     : PRIOR FILING DATE: 2001-01-30
14     : PRIOR APPLICATION NUMBER: PCT/US01/00661
15     : PRIOR FILING DATE: 2001-01-30
16     : PRIOR APPLICATION NUMBER: PCT/US01/00670
17     : PRIOR FILING DATE: 2001-01-30
18     : PRIOR APPLICATION NUMBER: US 60/234,687
19     : PRIOR FILING DATE: 2000-09-21
20     : PRIOR APPLICATION NUMBER: US 09/608,408
21     : PRIOR FILING DATE: 2000-06-30
22     : PRIOR APPLICATION NUMBER: US 09/774,203
23     : PRIOR FILING DATE: 2001-01-29
24     : NUMBER OF SEQ ID NOS: 49117
25     : SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
26     : SEQ ID NO: 1653
27     : LENGTH: 454
28     : TYPE: DNA
29     : ORGANISM: Homo sapiens
30     : FEATURE:
31     : OTHER INFORMATION: MAP TO AP000163.1
32     : OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.4
33     : OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.4
34     : OTHER INFORMATION: EXPRESSED IN HEPA, SIGNAL = 1.7
35     : OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.92
36     : OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.1
37     : OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.8
38     : OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.5
39     : OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.5
40     : OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1
41     : OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.95
42     : US-09-864-761-1653
43
44     Query Match      17.6%   Score 255;   DB 10;   Length 454;
45     Best Local Similarity 81.9%;   Pred. No. 3.7e-72;
46     Matches 294;   Conservative 0;   Mismatches 65;   Indels 0;   Gaps 0;
47
48     Oy 1060 ACTATGACAAATCTGACCGCTGACCTCGGTACTACTATGACAAATAATTATGACTAAAG 1119
49     Db 454 ACTACGATAGCTGACGCCGCCCTCCGTTACTATGACAAAGAACATCAATGACCAAG 395
50     Oy 1120 TTCATGGTAAAGCTATGCTCTACAAATTTGATTTTCACAGGAATGCTCAGGCCCTCCAGC 1179
51     Db 394 TCCATGGGAAGGCTACGCGCTACAAAGTTGCACTTCACAGGAGATCGCCACAGGCCCTCCAGC 335
52     Oy 1180 CTCACCCCTCCAGAAATCATCATGATGTCACAAATACCATACAGACTCCCTCATGAGTTCT 1239
53     Db 334 CCCACCCCCCGAGTCACTCTGTCAAGATGACCCCTCAGACCTCCCGTACAGGGGCTCCT 275
54     Oy 1240 ACCATGACACACCCCGAGAAGATGAACTTTGTAGCGCCCATCCCGCTGTTGGCCCGTAA 1299
55     Db 274 ATCACGCCACCCACAGAAAGATGAATCTGTGGGCCCCACCCCTCAGCGCTCCCGGTGA 215
56     Oy 1300 CCTCATCCAGCTTTTTCCTGCTGCCCTTAATTCATCTACTGGAATTCACCAACTGGAGGACTCT 1359
57     Db 214 CATCTCCAGTTTTCCTGCTGCCCCCAAAACCATACTGGAATTCACCAACTGGGGGTATAT 155
58     Oy 1360 ACCCAATACCAAGCTGCCAGAGTGTCTATATGCTTCATCTTGGCACTACTACTAA 1418
59     Db 154 ACCCAACACTAAGCTCCCAACAGCCATATGCTTCATCTGTGGCACTACTACTAA 96
60
61 RESULT 6
62 US-09-850-799-1
63 : Sequence 1, Application US/09850799

```


Query Match 11.3%; Score 163.2; DB 10; Length 477;
Best Local Similarity 86.5%; Pred. No. 1.7e-42;
Matches 180; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 448 CAGATCTACCTGTTATGACGACAGACATGACGAGTGTGCTGAGTGGGCGACTGAAG 507
DB 453 CAGATCTACCTGTTATGAGTACAGACATGTCGCGAGTGGTGGGCGGTGAAG 394
QY 508 AGTATGCTCTCCAGAGCTGACATCTTGTGTCCAGAAATGATGAGGAAGTGT 567
DB 393 AATATGGCTTCCAGAGCTCAACATCTTGTATTCAGAAATGATGAGGAAGTGT 334
QY 568 GTAAATGACCAAGATGACTTCCAGAGACTCACGCCGAGCTATPACCGACATATCTCC 627
DB 333 GCAAGATGACCAAGAGACTTCCAGAGGCTCACCCCGAGCTACAGCCGACATCTCTC 274
QY 628 TGTCACTCTACCTACTCTCAGAGAG 655
DB 273 TCTCACTCTCTCACTACTCTCAGAGAG 246

RESULT 10

US-09-864-761-12798/C
Sequence 12798, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aeomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263,6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117

SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 12798
LENGTH: 519
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AP001426.1
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.73
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.52
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.45
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.55
US-09-864-761-12798

Query Match 11.3%; Score 163.2; DB 10; Length 519;
Best Local Similarity 86.5%; Pred. No. 1.8e-42;
Matches 180; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 448 CAGATCTACCTGTTATGACGACAGACATGTAACGCGACTGCTGAGTGGGCGACTGAAG 507
DB 359 CAGATCTACCTGTTATGAGTACAGACATGTCGCGAGTGGTGGGCGGTGAAG 300
QY 508 AGTATGCTCTCCAGAGCTGACATCTTGTGTCCAGAAATGATGAGGAAGTGT 567
DB 299 AATATGGCTTCCAGAGCTCAACATCTTGTATTCAGAAATGATGAGGAAGTGT 240
QY 568 GTAAATGACCAAGATGACTTCCAGAGACTCACGCCGAGCTATPACCGACATATCTCTCC 627
DB 239 GCAAGATGACCAAGAGACTTCCAGAGGCTCACCCCGAGCTACAGCCGACATCTCTC 180
QY 628 TGTCACTCTACCTACTCTCAGAGAG 655
DB 179 TCTCACTCTCTCACTACTCTCAGAGAG 152

RESULT 11

US-09-864-761-21761/C
Sequence 21761, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
FILE REFERENCE: Aeomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263,6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662

```

RESULT 12
US-09-864-761-32394/C
Sequence 23294, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aemica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIORITY APPLICATION NUMBER: US 60/180,312
PRIORITY FILING DATE: 2000-02-04
PRIORITY APPLICATION NUMBER: US 60/207,456
PRIORITY FILING DATE: 2000-05-26
PRIORITY APPLICATION NUMBER: US 09/632,366
PRIORITY FILING DATE: 2000-08-03
PRIORITY APPLICATION NUMBER: GB 24263.6
PRIORITY FILING DATE: 2000-10-04
PRIORITY APPLICATION NUMBER: US 60/236,359
PRIORITY FILING DATE: 2000-09-27
PRIORITY APPLICATION NUMBER: PCT/US01/00666
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00667

```

QY	141	CACCTTCGCAAAACAGCAAAATACAGACCTCCCTCTTCCACAGTAATATGCGCAAAATATCAAG	201	ATGAGCCGCGCGCTTCCACAGCAGACTGGTATACAGAGCCCC	244
Db	165	GAAGCCTTATCAGTTGTGATGAGGAGACACGCTGTGTTGATGTGCTTACGGAAAGCCA	108		
QY	81	GAGACATTTATCAGTGTGATGAGGAAGACCACTCCTGTGTTGAGTGTGCTTACGAGTGGCC	144		
Db	45	ATGAGCCCCACGCGTCTCCACACAGGAATGGCTGTCTCAACCCC	2		
QY	201	ATGAGCCGCGCGCTTCCACAGCAGACTGGTATACAGAGCCCC	244		
Db	105	CACCTGCGCTTACAGCAGAAATATACCGCGTCTCTCTCCACGCACTATGAGACAGACTTCCAG	46		

RESULT 13
US-09-925-300-420
Sequence 420, Application US/09925300
Patent No. US20020151681A1
GENERAL INFORMATION:
APPLICANT: Craig Rosen,
APPLICANT: Steve Ruben
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA101
CURRENT APPLICATION NUMBER: US/09/925.300
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05988
PRIOR FILING DATE: 2000-03-08

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 9, 2002, 16:46:47 ; Search time 29.6136 Seconds

(without alignments)
1464.080 Million cell updates/sec

Title: US-09-902-772-2

Perfect score: 2444

Sequence: 1 MASTIKKALSVSEDSOLF.....IYNTRLPAAMPSHLGTYT 451

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2414.5	98.8	478	1	transcription fact
2	2298	94.0	486	1	transcription fact
3	1624.5	66.5	452	1	transforming prote
4	1613.5	66.0	452	1	transforming prote
5	1576	64.5	453	1	transforming prote
6	1375	56.3	385	1	transforming prote
7	1227	50.2	272	2	transcription fact
8	676	25.7	179	2	transcription fact
9	627.5	25.7	179	2	transcription fact
10	524	21.4	377	2	transcription fact
11	477	19.5	208	2	transcription fact
12	461	18.9	434	1	transcription fact
13	460	18.6	454	1	transcription fact
14	454.5	18.6	464	1	transcription fact
15	454	18.6	454	2	transcription fact
16	451.5	18.5	438	1	transcription fact
17	449	18.4	441	1	transcription fact
18	448	18.3	440	1	transcription fact
19	447	18.3	441	1	transcription fact
20	447	18.3	485	1	transcription fact
21	446	18.2	441	1	transcription fact
22	443.5	18.1	479	1	transcription fact
23	438	17.9	479	2	transcription fact
24	437.5	17.9	472	1	transcription fact
25	437	17.9	488	1	transcription fact
26	434.5	17.8	440	2	transcription fact
27	431.5	17.7	472	1	transcription fact
28	429	17.6	469	1	transcription fact
29	424.5	17.4	468	1	transcription fact

30	378	15.5	718	1	gene pointed prote
31	345	14.1	268	2	transcription fact
32	344.5	14.1	94	2	transcription fact
33	340.5	13.9	548	2	ETS2 repressor fac
34	338	13.8	211	2	transforming prote
35	336	13.7	623	1	gene pointed prote
36	321	13.1	110	2	transcription fact
37	320.5	13.1	477	1	transcription fact
38	318.5	13.0	250	2	transcription fact
39	314.5	12.9	477	1	transcription fact
40	311	12.7	393	2	transcription fact
41	305	12.5	732	2	transcription fact
42	300.5	12.3	510	1	transcription fact
43	298.5	12.2	342	2	transcription fact
44	286	11.7	555	1	transcription fact
45	285	11.7	428	1	transcription fact

ALIGNMENTS

RESULT 1

S60754

transcription factor erg - chicken

C:Species: Gallus gallus (chicken)

C:Date: 27-Apr-1996 #sequence_revision 23-May-1997 #text_change 16-Jul-1999

C:Accession: S60754

R:Diordain, P.; Dewitte, F.; Desbriens, X.; Stehelin, D.; Dutreque-Coquilaud, M.

Mech. Dev. 50, 17-28, 1995

A:Title: Mesodermal expression of the chicken erg gene associated with precartilagino

A:Reference number: S60754; MUID:95329425; PMID:7605748

A:Accession: S60754

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-478 <DHO>

A:Cross-references: EMBL:X77159; NID:9790439; PIDN:CA54404.1; PID:9790440

C:Genetics:

A:Gene: erg

C:Superfamily: transcription factor erg; ets DNA-binding domain homology; ets RII reg

C:Keywords: DNA binding; phosphoprotein; proto-oncogene; transcription factor; transf

F:119-193/Domain: ets RII regulatory region homology <ETR>

F:312-390/Domain: ets DNA-binding domain homology <ETS>

Query Match	98.8%	Score 2414.5	DB 1	Length 478
Best Local Similarity	94.1%	Pred. No. 3.2e-168		
Matches 450	Conservative	0	Mismatches 1	Indels 27
			Gaps	1
QY	1	MASTIKKALSVSEDSOLFECAYGSPHLAKTEMTASSSEYGTQSKMSRPVPODWSOP	60	
DB	1	MASTIKKALSVSEDSOLFECAYGSPHLAKTEMTASSSEYGTQSKMSRPVPODWSOP	60	
QY	61	PARTIKMECPNOVNGSRNSPDDCSYAGGKMTSSDNVGMNTGSYWEKHIPPNNTT	120	
DB	61	PARTIKMECPNOVNGSRNSPDDCSYAGGKMTSSDNVGMNTGSYWEKHIPPNNTT	120	
QY	121	NERRYVPADPTLWSTDHVQWLEMAVKEYGLPVVDILLFONIDGKELCKMTKDFORT	180	
DB	121	NERRYVPADPTLWSTDHVQWLEMAVKEYGLPVVDILLFONIDGKELCKMTKDFORT	180	
QY	181	PSYADILLSHLHYLRER-----GATFIPNTSVYPEA	213	
DB	181	PSYADILLSHLHYLRER-----GATFIPNTSVYPEA	213	
QY	214	TORTTRPDLPEQARRSAMTSHSHPTOSKATOPSSSTVPKTEPORPOLDYQILGPTSS	273	
DB	214	TORTTRPDLPEQARRSAMTSHSHPTOSKATOPSSSTVPKTEPORPOLDYQILGPTSS	273	
QY	241	TORTTRPDLPEQARRSAMTSHSHPTOSKATOPSSSTVPKTEPORPOLDYQILGPTSS	300	
DB	241	TORTTRPDLPEQARRSAMTSHSHPTOSKATOPSSSTVPKTEPORPOLDYQILGPTSS	300	
QY	274	RLANPGSQIQLMQFLLELLSDSSNSNCITWEGNGEKKMTDPDEVARMRERKSKPMN	333	
DB	274	RLANPGSQIQLMQFLLELLSDSSNSNCITWEGNGEKKMTDPDEVARMRERKSKPMN	333	
QY	301	RLANPGSQIQLMQFLLELLSDSSNSNCITWEGNGEKKMTDPDEVARMRERKSKPMN	360	
DB	301	RLANPGSQIQLMQFLLELLSDSSNSNCITWEGNGEKKMTDPDEVARMRERKSKPMN	360	
QY	334	YDKLSRALRYTDKNIMTKVAGKRAYKFDPHGIAQLQPPPPSSMTKTPYSSY	393	
DB	334	YDKLSRALRYTDKNIMTKVAGKRAYKFDPHGIAQLQPPPPSSMTKTPYSSY	393	

DB 361 YDKLSALRLRYDYDKNMTYHGKRYAKFDFHGIADLPHPRESSMYKPSDLPYMSY 420

QY 394 HAHPOKMNFAHPHPALPYTSSSFFAAPNPYMSPTGGIYPTNRLPAAHMPSHLGTY 451
 |||||||

DB 421 HAHPOKMNFAHPHPALPYTSSSFFAAPNPYMSPTGGIYPTNRLPAAHMPSHLGTY 478
 |||||||

RESULT 2

TRANSFORMING PROTEIN ERG-3 - human

N:Contains: transforming protein erg-1; transforming protein erg-2

C:Species: Homo sapiens (man)

C:Date: 31-Mar-1989 #sequence_revision 30-May-1997 #text_change 16-Jul-1999

C:Accession: A94294; A94178; I58410; A28041; A29515

R:Rao, V.N.; Papas, T.S.; Shyam, E.; Reddy, P.

S:Science 237, 635-639, 1987

A:Title: erg, a human ets-related gene on chromosome 21: alternative splicing, polyadeny

A:Reference number: A94294; MUID:87263429; PMID:3299708

A:Accession: A94294

A:Molecule type: mRNA

A:Residues: 1-231,256-486 <RE>

A:Cross-references: GB:M17254; NID:g182186; PIDN:AAA52398.1; PID:g182187

R:Reddy, E.S.P.; Rao, V.N.; Papas, T.S.

Proc. Natl. Acad. Sci. U.S.A. 84, 6131-6135, 1987

A:Title: The erg gene: a human gene related to the ets oncogene.

A:Reference number: A94178; MUID:87317608; PMID:3476534

A:Accession: A94178

A:Molecule type: mRNA

A:Residues: 100-231,256-486 <RE>

A:Cross-references: GB:M21535; NID:g182182; PIDN:AAA52811.1; PID:g182185

R:Prasad, D.D.; Rao, V.N.; Lee, L.; Reddy, E.S.

Oncogene 9, 669-673, 1994

A:Title: Differentially spliced erg-3 product functions as a transcriptional activator.

A:Reference number: I58410; MUID:94119611; PMID:8290279

A:Accession: I58410

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 230-259 <RE>

A:Cross-references: GB:S68130; NID:g544960; PIDN:AA29724.1; PID:g544961

C:Genetics: GDB: GDB:ERG

A:Gene: GDB:ERG

A:Cross-references: GDB:119884; OMIM:165080

A:Map position: 21q22.2-21q22.2

C:Superfamily: transcription factor erg; ets DNA-binding domain homology; ets RII regula

C:Keywords: alternative splicing; DNA binding; nucleus; proto-oncogene; transcription fa

F:1-231,256-486/Product: transforming protein erg-2 #status predicted <K&G>

F:100-231,256-486/Product: transforming protein erg-1 #status predicted <EG>

F:126-200/Domain: ets RII regulatory region homology <EMR>

F:320-398/Domain: ets DNA-binding domain homology <EMR>

Query Match 94.0%; Score 2298; DB 1: Length 486;
 Best Local Similarity 89.1%; Pred. No. 1e-159;
 Matches 426; Conservative 10; Mismatches 14; Indels 28; Gaps 2;

QY 2 ASITKALSVSDOSLFCAGSPHLAKTEMASSSSSEYSGSKSPVPOODWLSOPP 61
 |:|||||

DB 9 AAHKAALSVSDOSLFCAGSPHLAKTEMASSSSSTGGTSKSPVPOODWLSOPP 68
 |||||||

QY 62 ARVTKMECPNPNVNGSRNSPDDCSYAKGKMYSSSDNGMYSGYMEKHIPPMMNTN 121
 |||||||

DB 69 ARTIMECPNPNVNGSRNSPDDCSYAKGKMYSPDYGMNYSMEKHPMPMMNTN 128
 |||||||

QY 122 ERRVVPAPPTLMSIDHVRQWLEMAVKEYGLDPVDILLFQNDIGKELCKMTKDDFORLTP 181
 |||||||

DB 129 ERRVVPAPPTLMSIDHVRQWLEMAVKEYGLDPVDILLFQNDIGKELCKMTKDDFORLTP 188
 |||||||

QY 182 SYNADILLSHLYLRER-----GATTFPPTSYVPEAT 214
 |||||||

DB 189 SYNADILLSHLYLRERPHLTSDVDKALONSPLMHARNTGGAATFPPTSYVPEAT 248
 |||||||

QY 215 QITTRPDLPEEQARRSAMTSHSHP-OSKATOPSSSYTKTEDORPOLDPQOILGPTSS 273
 |||||||

DB 249 QITTRPDLPEEQARRSAMTSHSHP-OSKATOPSSSYTKTEDORPOLDPQOILGPTSS 308
 |||||||

QY 274 RLAPSGQIQLMQFLLELLSDSSNSNCITWECTNGEFKWTDEVARRRGERRSKPNMN 333
 |||||||

DB 309 RLAPSGQIQLMQFLLELLSDSSNSNCITWECTNGEFKWTDEVARRRGERRSKPNMN 368
 |||||||

QY 334 YDKLSALRLRYDYDKNMTYHGKRYAKFDFHGIADLPHPRESSMYKPSDLPYMSY 393
 |||||||

DB 369 YDKLSALRLRYDYDKNMTYHGKRYAKFDFHGIADLPHPRESSMYKPSDLPYMSY 428
 |||||||

QY 394 HAHPOKMNFAHPHPALPYTSSSFFAAPNPYMSPTGGIYPTNRLPAAHMPSHLGTY 451
 |||||||

DB 429 HAHPOKMNFAHPHPALPYTSSSFFAAPNPYMSPTGGIYPTNRLPAAHMPSHLGTY 486
 |||||||

RESULT 3

TRANSFORMING PROTEIN FLI, long splice form - human

N:Alternate names: Friend leukemia integration protein 1; transcription factor ERGB

C:Species: Homo sapiens (man)

C:Date: 04-Oct-1996 #sequence_revision 30-May-1997 #text_change 03-Mar-2000

C:Accession: I37565; S29843; S35506; A49000; A49015; I54170

R:Delattre, O.; Zucman, J.; Plougastel, B.; Desmaza, C.; Melot, T.; Peter, M.; Kovar, Nature 359, 162-165, 1992

A:Title: Gene fusion with an ETS DNA-binding domain caused by chromosome translocatio

A:Reference number: S28257; MUID:92396239; PMID:1522903

A:Accession: I37565

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-452 <DE>

A:Cross-references: EMBL:670001; NID:g32529; PIDN:CAA47399.1; PID:g32530

R:Hromas, R.; May, W.; Denny, C.; Raskind, W.; Moore, J.; Makl, R.A.; Beck, E.; Klems Blochm, Biophys. Acta 1172, 155-158, 1993

A:Title: Human FLI-1 localizes to chromosome 11Q24 and has an aberrant transcript in

A:Reference number: S29843; MUID:93176799; PMID:8439553

A:Accession: S29843

A:Molecule type: mRNA

A:Residues: 1-68, 'V', '70-129, 'A', '131-132, 'V', '134-322, 'Q', '324-425, 427-452 <HRO>

A:Cross-references: EMBL:M93255; NID:g182659

A:Note: the authors translated the codon GTC for residue 69 as Glu

R:Hromas, R.A.; May, W.; Denny, C.; Raskind, W.; Moore, J.; Makl, R.A.; Beck, E.; Kle submitted to the EMBL Data Library, May 1993

A:Description: Human FLI-1, an ETS oncogene family member preferentially expressed in

A:Reference number: S35506

A:Accession: S35506

A:Molecule type: mRNA

A:Residues: 1-68, 'V', '70-76, '78-129, 'A', '131-132, 'V', '134-322, 'Q', '324-425, 427-452 <HRO>

A:Cross-references: GB:S45205; NID:g257353; PIDN:AA23637.1; PID:g257354

A:Note: sequence extracted from NCBI Backbone (NCBIN:115336, NCBI:P115337)

R:Watson, D.K.; Smyth, F.E.; Thompson, D.M.; Cheng, J.Q.; Testa, J.R.; Paps, T.S.; S Cell Growth Differ. 3, 705-713, 1992

A:Title: The ERGB/FLI-1 gene: isolation and characterization of a new member of the f

A:Reference number: A49015; MUID:93075640; PMID:1445800

A:Accession: A49015

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-390, 'Q', '392-425, 427-452 <WAT>

A:Molecule type: mRNA

A:Cross-references: GB:M98833; NID:g182188; PIDN:AAA52812.1; PID:g182188

A:Note: sequence extracted from NCBI Backbone (NCBIN:119390, NCBI:P119392)

R:Band, V.; Lipinski, M.; Raastart, E.; Polliquin, L.; Bergeron, D.

Genomics 11, 223-224, 1991

A:Title: The human homolog of the mouse common viral integration region, FLI1, maps t

A:Reference number: I54170; MUID:92112219; PMID:1765382

A:Accession: I54170

A:Status: preliminary; nucleic acid sequence not shown; translated from GB/EMBL/DBJ

A:Molecule type: DNA
A:Residues: 1-6 <BAD>
A:Cross-references: GB:147616; NID:g1000863; PIDN:AAA6854.1; PID:g1000864
C:Genetics:
A:Gene: GDB:FLI1
A:Cross-references: GDB:127565; OMIM:193067
A:Map position: 11q24.1-11q24.3
A:Introns: 77/3
C:Superfamily: transcription factor ery; ets DNA-binding domain homology; ets R11 regulatory factor 1; ets R11 regulatory region homology <ETR>
C:Keywords: alternative splicing; DNA binding; nucleus; proto-oncogene; transcription factor
F:118-192/Domain: ets R11 regulatory region homology <ETR>
F:283-361/Domain: ets DNA-binding domain homology <ETS>

	Score	DB 1	Length
Query Match	66.5%	1624.5	452
Best Local Similarity	67.0%	1.1e-110	
Matches 306; Conservative	57;	Mismatches 83;	Indels 11; Gaps 7;

```

QY 1 MASUKKELSVYSDOLFECAYG - SPHLAKTMEHTASSSSEYQOTSKMSRPVQDWMQ 59
Db 1 MGGTKEKELSVSDQSDJFDSATGAAHLEKRAKMDJASGSDPYQDPKHINLPPEQWEMIO 60
QY 60 PPARVYTKMECNPQVNGSRNSPDDCSVAKGGRKMWSSSDVGNAGNYSYMEEKH - IPPNM 118
Db 61 -PVRANYKREX - DHMNGSRSPYDGSVSKCSKLTGVGGESNPANNYSYNDENKGRPPNM 117
QY 119 TTNERRYVVPADPLTJNSTDHYRQMLEVAAYEYGLPDVDILLFONIDGKELCKMTKDDFOR 178
Db 118 TTNRERRVVPADPLTJMOEHYRQMLEWAIKEYSIMEJLDTSFONMKGKELCKKNNKEDFJR 177
QY 179 LTPSYNDILLSHTHYRREGKATFIPNTSVYPEAOTIRITRBDPLDYEQARSGAMTSHS 238
Db 178 ATTLXNEVLLSHSLTYRE - SSILAVNTTSHDQSSRLSKEDPSYDSYRGAMGNMNN 235
QY 239 PTQSKAQO - PSSSTVPYTEDQRPQLDPYQILGPTSRKANPNSGGQIQLOMFLLELLSDS 297
Db 236 SGLKMSPLIGAQOTISKTEBQRPQDPDYQILGPTSRKANPNSGGQIQLOMFLLELLSDSA 295
QY 298 NSNOCITTEGTGGEKMTDPEVARRMGEKSKPMNNDKLSRLRYRYOKNITKTHGR 357
Db 296 NASCTWGTGNGEKKMTDPEVARRMGEKSKPMNNDKLSRLRYRYOKNITKTHGR 355
QY 358 YAYFDFEFGIQAOLQPHRPRESSMKYPSDLPYSSYHAHPEQKANFAPHPALPYSSSF 417
Db 356 YAYKFDFEFGIQAOLQPHRPTESSMKYPSDLSYMPYHAHQKYNFPHRPSMPYSSSF 415
QY 418 FAPNPPTWNSPTGIGYFN --TRLPAAHMPSHJGTY 451
Db 416 FGAASQYWTSEPTGIGYFNPNVPHRPNTHVSHSGSY 452

```

RESULT 4
S17403

N:Alternate names: Friend leukemia integration protein 1; transcription factor ERGB
C:Species: Mus musculus (house mouse)

R. Ben-David, Y.; Giddens, E.B.; Letwin, K.; Bernstein, A. *Genes Dev.* 5, 908-918, 1991

A: Molecule type: mRNA
A: Accession: SI7403
Reference Number: SI7403, MOLD:SI2370, PMID:404553

A:cross-references: GB:A59421; NID:g509/4; PIDN:CAA42055.1; PID:g509/5
C:Genetics:
A:Gene: F11

F.118-192/Domain: ets RII regulatory region homology <ETR>
F.283-361/Domain: ets DNA-binding domain homology <ETS>

Query Match	66.08; Score 1613.5; DB 1; Length 452;
-------------	--

Best Local Similarity 66.7%; Pred. No. 6,7e-110;
Matches 305; Conservative 56; Mismatches 85; Indels 11; Gaps 7.

Y 1 MASTIKELSVSEDOLEFCAYG-SPHLAKTEMTASSSSSEYGQTSKMSPRVPQQDWLSQ 59

b 1 MDGTIKEALSVSDQSLFDSA YGA AHL PKADMTASGSPDYGQPHKINPLPQQEWINQ 60

60 PPARVTIKMECNPNQVNGSRNSPDDCSVAKGKMWSSSDNVGMNYGSYMEEKH-IPPPNM 118

b 61 -PVRVNVKREY - -DHMGSRSPVDCSVSKCNKLVGGGEANPMNYSYMDKNGPPPNM 117

119 TTNERRVIVPADPTLMSTDHVRQWLEWAVKEYGLPDVDILLFQNIIDGKELCKMTKDDFOR 178

118 T T N E R R V I P A D P T L M T Q E H V R Q L E M A I K E Y G L M E I D T S F Q N M D G K E L C M K N K E D F L R 177

1/9 LIPSYNADILLSHLHYLRERGATELEPNISVYPEAIIQRIITRPDLPYEQARRSAMISSH 238

170 A15A1N1EVL5H51 LK5 - 55LH4N1 1SH1D55K5N5K5E5D51D55V5K5G5A5N5N5M5N 235

[illegible]

U
C
B
Z
U
B
C
U
K
B
C
Z
B
K
K
U
K
B
C
C
C
K
K
B
Z
K
B
B
C
C
C
B
B
C
U

[illegible]

.....

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

[illegible]

```

0      | : | | | | | | | | | | | | | |
415    ECASCVWTSPTACTYBNDSCVBBDNMTTHUCUICGVV   A53

```

RESULT 5
49013

transferring protein III - African clawed frog

```

/species: xenopus laevis (african clawed frog)
Date: 07-May-1995 #sequence_revision 23-May-1997 #text_change 16-Jul-1999

```

/accession: 042013
Meyer, D.; Wolff, C.M.; Stiegler, P.; Senan, F.; Befort, N.; Befort, J.J.; Re

Title: *Xl-fli*, the *Xenopus* homologue of the *fli-1* gene, is expressed during e

Accession: S49013

; molecule type: mRNA

;Cross-references: EMBL:X66979; NID:g505486; PIDN:CAA47389.1; PID:g505487

transcription factor encoding DNA-binding domain and a

Keywords: DNA binding; nucleus; proto-oncogene; transcription factor; transfection; KeyWords: DNA binding; nucleus; proto-oncogene; transcription factor; transfection

;284-362/Domain: ets DNA-binding domain homology <ETS>

query match	84.38;	score 13/6;	DB 1;	length 453;
Best Local Similarity	66.28;	Pred. No. 3.6e-107;		

Year	Conservative	Misadventures	Standards	Gaps
1996	100	100	100	100
2000	100	100	100	100
2004	100	100	100	100
2008	100	100	100	100
2012	100	100	100	100
2016	100	100	100	100
2020	100	100	100	100
2024	100	100	100	100
2028	100	100	100	100
2032	100	100	100	100
2036	100	100	100	100
2040	100	100	100	100
2044	100	100	100	100
2048	100	100	100	100
2052	100	100	100	100
2056	100	100	100	100
2060	100	100	100	100
2064	100	100	100	100
2068	100	100	100	100
2072	100	100	100	100
2076	100	100	100	100
2080	100	100	100	100
2084	100	100	100	100
2088	100	100	100	100
2092	100	100	100	100
2096	100	100	100	100
2100	100	100	100	100

[illegible]

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

RESEARCH DESIGN AND METHODS

||||| :||| :|| | ||||| :|||

Db 117 TTRERVIYPADALMSODHVRQMLEAKKEGLVEIDCSLFONIDKELCKSKDEFLR 176
 QY 179 LPESYNADILLSHLYLRERGAFTFEPNTSVYPEATQRTTRPDLPEQARRSAM-TSHS 237
 Db 177 STSIYMTVEVLLSHLNTLRDSSSLG-NTQAHNTDQSSRLTAKEDPSEAVRSGWGSMS 235
 QY 238 HPIQSKATQPSSTVKTED-QRPQDLPYQILGPTSSRLANPGSGQIQLMQFLELLSDS 236
 Db 236 SPYTKSPPMGGTQNVKSGQOQRPDPYQILGPTSSRLANPGSGQIQLMQFLELLSDS 295
 QY 297 SNSNCITTEGTCGFMTPDEVARRMGERKSPNNMYDKLSRLRYDDKNITKTVHGK 356
 Db 296 SNSNCITTEGTCGFMTPDEVARRMGERKSPNNMYDKLSRLRYDDKNITKTVHGK 355
 QY 357 RYAKFDFHGAIALQPHPESSMYKPSDLPYSSYHAHPQKMFVAPHPALPTSSS 416
 Db 356 RYAKFDFHGAIALQPHPESSMYKPSDLPYSSYHAHPQKMFVAPHPALPTSSS 415
 QY 417 FFAAPRPYNNSPITGILYRN---TRLPAAHMPSHLGITY 451
 Db 416 FFAAPRPYNNSPITGILYRN---TRLPAAHMPSHLGITY 453

RESULT 6

S29844
 Transforming protein flt, short splice form - human
 M:Alternate names: Friend leukemia integration protein 1; transcription factor ERGB
 C:Species: Homo sapiens (man)
 C>Date: 02-Dec-1993 #sequence_revision 23-May-1997 #text_change 16-Jul-1999
 C:Accession: S29844
 R:Hromas, R.; May, W.; Denny, C.; Raskind, W.; Moore, J.; Maki, R.A.; Beck, E.; Klametz,
 Blochm. Biophys. Acta 1172, 155-158, 1993
 A:Title: Human FLT-1 localizes to chromosome 11Q24 and has an aberrant transcript in neu
 A:Reference number: S29843; MUID:93176799; PMID:8439553
 A:Accession: S29844
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: mRNA
 A:Residues: 1-385 <RQ>
 A:Cross-References: EMBL:M93255; NID:g18265; PIDN:AA58480.1; PID:g182661
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1992
 C:Genetics:
 A:Gene: GDB:FLI1
 A:Cross-References: GDB:127565; OMIM:193067
 A:Map position: 11q24.1-11q24.3
 C:Superfamily: transcription factor erg; ets DNA-binding domain homology; ets RII regula
 C:Keywords: alternative splicing; DNA binding; nucleus; proto-oncogene; transcription fa
 F:52-126/Domain: ets RII regulatory region homology <Erg>
 F:217-295/Domain: ets DNA-binding domain homology <Erg>

Query Match 56.3%; Score 1375; DB 1; Length 385;

Best Local Similarity 67.8%; Pred. No. 1.3e-92;

Matches 257; Conservative 46; Mismatches 68; Indels 8; Gaps 5;

QY 78 SRNSPDCSVAGKGMVSSSDNYGMNGYSMEKH-IPPNNMTNRRYIVPADPLWST 136
 Db 10 ARESPYDCSVSKSLVGGESPMNPNYNSYDEKNGPPPNMTNRRYIVPADPLWST 69
 QY 137 DHYRQMLEAVYEGLPDVLFFONIDKELCKMTKDDFORLTPSYNADILLSHLYLR 196
 Db 70 EHVHQLLEMAIKESYLMETDTSFQMDKELCKMKEDFLATITLYNEVLLSHLYLR 129
 QY 197 EKGATFIFNTSYVPEATQRTTRPDLPEQARRSAMTSHSHPTQSAQO-PSSSTVPT 255
 Db 130 E-SSLAANTSTHTQSSRLSYKEDPSYSVRGAMGNMNSGLKSPPLGAQYISKN 187
 QY 256 EDOHPOLDYQILGPTSSRLANPGSGQIQLMQFLELLSDSSNSNCITTEGTCGFM 315
 Db 188 TEORPQDPDYQILGPTSSRLANPGSGQIQLMQFLELLSDSSNSNCITTEGTCGFM 247
 QY 316 PDEVARRMGERKSPNNMYDKLSRLRYDDKNITKTVHGKRAYFDFHGAIALQPHR 375
 Db 248 PDEVARRMGERKSPNNMYDKLSRLRYDDKNITKTVHGKRAYFDFHGAIALQPHR 307

QY 376 PESSMYKPSDLPYSSYHAHPQKMFVAPHPALPTSSSFFAADPNYNNSPITGILYRN 435
 Db 308 TESSMYKPSDLPYSSYHAHPQKMFVAPHPALPTSSSFFAADPNYNNSPITGILYRN 366
 QY 436 ---TRLPAAHMPSHLGITY 451
 Db 367 PNVPRHNPVSHLGSY 385

RESULT 7

A54617
 transcription factor erg - mouse (fragment)
 M:Alternate names: Ig heavy chain enhancer-binding protein
 C:Species: Mus musculus (house mouse)
 C>Date: 25-Oct-1994 #sequence_revision 18-Nov-1994 #text_change 16-Jul-1999
 C:Accession: A54617
 R:Rivera, R.R.; Stulver, M.H.; Steenbergen, R.; Murte, C.
 Mol. Cell. Biol. 13, 7163-7169, 1993
 A:Title: Ets proteins: new factors that regulate immunoglobulin heavy-chain gene expr
 A:Reference number: A54617; MUID:94019387; PMID:8413305
 A:Accession: A54617
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-272 <RI>
 A:Cross-References: GB:S6169; NID:g436276; PIDN:AA28525.1; PID:g436277
 A:Experimental source: pre-B-cell line 22D6
 A:Note: sequence extracted from NCBI backbone (NCBIN:138523, NCBI:P138524)
 C:Superfamily: transcription factor erg; ets DNA-binding domain homology; ets RII reg
 C:Keywords: DNA binding; phosphoprotein; proto-oncogene; transcription factor; transf
 F:140-218/Domain: ets DNA-binding domain homology <Erg>

Query Match 50.2%; Score 1227; DB 2; Length 272;

Best Local Similarity 85.3%; Pred. No. 4.7e-82;

Matches 232; Conservative 2; Mismatches 10; Indels 28; Gaps 2;

QY 174 DDFQRLTPSYNADILLSHLYLRER-----GATFIFPN 206
 Db 1 DDFQRLTPSYNADILLSHLYLRERPLPHTLSDVDKALONSPRLMHAARNTGGAFFPN 60
 QY 207 TSYVPEATQRTTRPDLPEQARRSAMTSHSHPT-OSKATOPSSSTVPTEDQRPOLDPY 265
 Db 61 TSYVPEATQRTTRPDLPEQARRSAMTSHSHPT-OSKATOPSSSTVPTEDQRPOLDPY 120
 QY 266 QILGPTSSRLANPGSGQIQLMQFLELLSDSSNSNCITTEGTCGFMTPDEVARRMGE 325
 Db 121 QILGPTSSRLANPGSGQIQLMQFLELLSDSSNSNCITTEGTCGFMTPDEVARRMGE 180
 QY 326 RRSKPMNNTDKLSRLRYDDKNITKTVHGKRAYFDFHGAIALQPHPESSLYKYP 385
 Db 181 RRSKPMNNTDKLSRLRYDDKNITKTVHGKRAYFDFHGAIALQPHPESSLYKYP 240
 QY 386 DLPEYSSYHAHPQKMFVAPHPALPTSSSF 417
 Db 241 DLPEYSSYHAHPQKMFVAPHPALPTSSSF 272

RESULT 8

A56646
 transcription factor erg/Flt-1 homolog - sea urchin (Lytechinus variegatus) (fragment)
 C:Species: Lytechinus variegatus (variegated urchin)
 C>Date: 03-Oct-1995 #sequence_revision 03-Oct-1995 #text_change 16-Jul-1999
 C:Accession: A56646
 R:Qi, S.; Chen, Z.Q.; Papas, T.S.; Lautenberger, J.A.
 DNA Seq. 3, 127-130, 1992
 A:Title: The sea urchin erg homolog defines a highly conserved erg-specific domain.
 A:Reference number: A56646; MUID:93091246; PMID:1457815
 A:Accession: A56646
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-173 <QIA>
 A:Cross-References: GB:M81067; NID:g161310; PIDN:AA68905.1; PID:g161311
 C:Superfamily: transcription factor erg; ets DNA-binding domain homology; ets RII reg
 C:Keywords: DNA binding; nucleus; transcription factor

F:6-84/Domain: ets DNA-binding domain homology <ETS>

Query Match 27.7%; Score 676; DB 2; Length 173;

Best Local Similarity 73.3%; Pred. No. 3.5e-42;

Matches 129; Conservative 16; Mismatches 23; Indels 8; Gaps 4;

QY 280 SGOIQIOMQFLELLSDSSNSNCITWEGTNGEFGKMTDPEVARMRGKSKPMNMYDKLSR 339

DB 1 SGOIQIOMQFLELLSDSSNSNCITWEGTNGEFGKMTDPEVARMRGKSKPMNMYDKLSR 60

QY 340 ALRYYYDKNIMTKVHGKRYAKFPDFHGAIALQPHPPSSMYKYPSPDLPYMSYHAHPQK 399

DB 61 ALRYYYDKNIMTKVHGKRYAKFPDFHGAIALQPHPPSSMYKYPSPDLPYMSYHAHPQK 118

QY 400 MNEFVAPHPALPYTSSSFFAAPNPYNSPTG-GIYPN---TRLPAHPSHLGTY 451

DB 119 LNFVG--TPINPSTNASLFSHSSSYMSPTGANIYPSGHVYHPASHMSHIGTY 172

RESULT 9

S51226 transcription factor erg/fli-1 homolog - polychaete (Nereis diversicolor) (fragment)

C:Species: Nereis diversicolor (sandworm)

C>Date: 15-Jul-1995 #sequence_revision 23-May-1997 #text_change 30-May-1997

C:Accession: S51226

R:Leleuvre-Cholteau, A.; Laudet, V.; Flourrens, A.; Begue, A.; Lepince, D.; Fontaine, F.

FEBS Lett. 354, 62-66, 1994

A:Title: Identification of two ets related genes in a marine worm, the polychaete annelid

A:Reference number: S51226; MUID:55046307; PMID:7957902

A:Accession: S51226

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-179 <LEL>

C:Superfamily: transcription factor erg; ets DNA-binding domain homology; ets RII regula

C:Keywords: DNA binding; nucleus; transcription factor

F:7-85/Domain: ets DNA-binding domain homology <ETS>

Query Match 25.7%; Score 627.5; DB 2; Length 179;

Best Local Similarity 68.2%; Pred. No. 1.3e-38;

Matches 122; Conservative 15; Mismatches 33; Indels 9; Gaps 5;

QY 279 GSGGIOMQFLELLSDSSNSNCITWEGTNGEFGKMTDPEVARMRGKSKPMNMYDKLS 338

DB 1 GSGGIOMQFLELLSDSSNSNCITWEGTNGEFGKMTDPEVARMRGKSKPMNMYDKLS 60

QY 339 RALRYYYDKNIMTKVHGKRYAKFPDFHGAIALQPHPPSSMYKYPSPDLPYMSYHAHPQ 398

DB 61 RALRYYYDKNIMTKVHGKRYAKFPDFHGAIALQPHPPSSMYKYPSPDLPYMSYHAHPQ 118

QY 399 KMFVAPHPALPYTSSSFFAAPNPYNSPTG-GIYPNTRLPA-----AHMPSHLGTY 451

DB 119 KLMMAAHAP-MASSASGFFPPAPYMSLGVSNLYPNISNHAHSHHGHMSHLGSY 176

RESULT 10

T34343 hypothetical protein T08H4.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 02-Sep-2000

C:Accession: T34343

A:Status: preliminary; translated from GB/EMBL/DBD

A:Molecule type: DNA

A:Residues: 1-377 <STE>

A:Cross-references: EMBL:U09470; PIDN:AACT1120.1; GSPDB:GN00020; CESP:T08H4.3

C:Genetics:

A:Gene: CESP:T08H4.3

A:Map position: 2

A:introns: 63/3; 88/1; 111/3; 146/1; 194/1; 297/1; 343/3

C:Species: ets DNA-binding domain homology

F:216-294/Domain: ets DNA-binding domain homology <ETS>

Query Match 21.4%; Score 524; DB 2; Length 377;

Best Local Similarity 51.9%; Pred. No. 1.2e-30;

Matches 120; Conservative 14; Mismatches 30; Indels 67; Gaps 10;

QY 226 EOARRSAMTSYSH-----PTQSK-----ATQPSSTVPTED 257

DB 110 DQSRQRYTESNSNGAATSSGSSSTEEKSDVFNISMNAPATPS-----KSED 165

QY 258 QR-POL-----DPYQILGPTSSRLANPGSGQIQIOMQFLELLS 294

DB 166 HNIPSFNMLSSYYTGALKLSNSTSFANPDYQILGPTSKNLHSGSCQIQIOMQFLELLS 225

QY 295 DSSNSNCITWEGTNGEFGKMTDPEVARMRGKSKPMNMYDKLSRALRYYYDKNIMTKVH 354

DB 226 DKRYSEVITWEGTNGEFGKMTDPEVARMRGKSKPMNMYDKLSRALRYYYDKNIMTKVH 285

QY 355 GKRYAKFPDFHGAIALQPHPPSSMYKYPSPDLPYMSYHAHPQKMFVAP 405

DB 286 GKRYAKFPDFHGAIALQ---PPTAS---HPQD---YFNS-HA---MGRIP 324

RESULT 11

T15790 hypothetical protein C42D8.4 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 02-Sep-2000

C:Accession: T15790

R:Hallsworth, K.

A:Title: The sequence of C. elegans cosmid C42D8.

A:Reference number: Z18405

A:Accession: T15790

A:Status: preliminary; translated from GB/EMBL/DBD

A:Molecule type: DNA

A:Residues: 1-208 <HAL>

A:Cross-references: EMBL:U05966; NID:q1293844; PID:q1293845; PIDN:AAA98717.1; GSPDB:G

A:Experimental source: strain Bristol N2; clone C42D8

C:Genetics:

A:Gene: CESP:C42D8.4

A:Map position: X

A:introns: 80/2; 104/3; 151/1; 187/3

C:Superfamily: ets DNA-binding domain homology

F:37-115/Domain: ets DNA-binding domain homology <ETS>

Query Match 19.5%; Score 477; DB 2; Length 208;

Best Local Similarity 46.6%; Pred. No. 1.4e-27;

Matches 103; Conservative 29; Mismatches 41; Indels 48; Gaps 6;

QY 247 PSSSTVKTEDQRPQDLPYQILGPTSSRLANPGSGQIQIOMQFLELLSDSSNSNCITWEG 306

DB 11 PSSNSTASK-----GPMALALSATGTGQIQIOMQFLELLADAVNAHCIAWEG 58

QY 307 TNGEFGKMTDPEVARMRGKSKPMNMYDKLSRALRYYYDKNIMTKVHGKRYAKFPFG 366

DB 59 SNGEFGKMTDPEVARMRGKSKPMNMYDKLSRALRYYYDKNIMTKVHGKRYAKFPFG 118

QY 367 IAQALQ-----PHPPSSMYKYPSPDLPY-----MSSYHAHPQK 401

DB 119 LAQCQSAIITLNGCNPNGLDLSYVHSLS--PYTQVLPITGTSRLSTSMSSYHLSLSTS 176

QY 402 FVAPHPALPYTSSSFFAAPNPYNSPTG-GIYPNTRLPAH 442

DB 177 STSSNQIIPSTAT-----YWTQSSSL-TYTGMPSSY 208

RESULT 12

A40858 GA-binding protein alpha chain - mouse

N:Alternate names: GABPA; nuclear respiratory factor-2 alpha chain; transcription fac

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 9, 2002, 16:55:42 ; Search time 2920.58 Seconds

(without alignments)
4494.100 Million cell updates/sec

Title: US-09-902-772-2

Perfect score: 2444
Sequence: 1 MASTIKKALSVSEDSLEF.....IYPTRLPAHMSHGTYY 451

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xip
-O=/cgn2_1/USPTO.spool/US09902772/runat.07112002.134033.9031/app.query.fasta_1.1294
-DB=GenDbml -GFWT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blonsum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR_SCORE=pcr -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pic -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09902772.gcg.1.1.3441.@runat.07112002.134033.9031 -NCPU=6 -ICPU=3
-NO_XIPYX -NO_MMAR -LARGEOURRY -NEG_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-WARN.TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenDbml: *
1: gb_ba: *
2: gb_hlg: *
3: gb_in: *
4: gb_om: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sts: *
12: gb_sy: *
13: gb_un: *
14: gb_vi: *
15: em_ba: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_mu: *
20: em_om: *
21: em_or: *
22: em_ov: *
23: em_pat: *
24: em_ph: *
25: em_pl: *
26: em_ro: *
27: em_sts: *
28: em_un: *

29: em_vi: *
30: em_htg_hum: *
31: em_htg_inv: *
32: em_htg_other: *
33: em_htg_mus: *
34: em_htg_pln: *
35: em_htg_rtd: *
36: em_htg_mam: *
37: em_htg_vrt: *
38: em_sy: *
39: em_htgo_hum: *
40: em_htgo_mus: *
41: em_htgo_other: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	2444	100.0	1413	5 AY065661	AY065661 Gallus ga
2	2444	100.0	1447	6 E31253	E31253 Proteins hav
3	2414.5	98.8	1516	5 G6ERG	X77159 G.gallus ER
4	2414.5	98.8	1528	6 E31254	E31254 Proteins hav
5	2267	92.8	2209	10 AB073080	AB073080 Mus muscu
6	2192	89.7	2012	5 XLAJ4126	AJ224126 Xenopus l
7	2182	89.3	3166	9 HOMERG2	AJ224126 Xenopus l
8	2179	89.2	1808	10 AB031088	M17254 Human erg2
9	2159.5	88.4	2133	10 AB073079	AB073079 Mus muscu
10	2153.5	88.1	1798	5 XLAJ4125	AB073079 Mus muscu
11	2151	88.0	2131	10 AB073078	AJ224125 Xenopus l
12	1810	74.1	3126	5 HOMERG11	AB073078 Mus muscu
13	1637	67.0	1436	5 DRE249550	M21535 Human erg p
14	1624.5	66.5	1359	9 AY029368	AJ249590 Danilo rer
15	1624.5	66.5	2394	9 BC010115	AY029368 Homo sapi
16	1624.5	66.5	2916	6 A36461	BC010115 Homo sapi
17	1624.5	66.5	2938	6 AR080101	A36461 Sequence 2
18	1624.5	66.5	2938	6 HSHMFL1	AR080101 Sequence
19	1624.5	66.5	2957	9 HOMERGBFLI	X67001 H.sapiens H
20	1624.5	66.5	2959	9 BC001670	M88833 Homo sapien
21	1624.5	66.5	6849	12 AY029367	BC001670 Homo sapi
22	1621.5	66.3	1673	9 S45205	AY029367 Synthetic
23	1613.5	66.0	1729	10 MMFLI1	S45205 Flt-1-Frien
24	1579.5	64.6	3490	5 CCRNAFLI	X59421 Mouse Flt-1
25	1576	64.5	4403	5 XIFLIG	Y14773 Coturnix co
26	1567.5	64.1	3545	5 CCEFLONCO	X66979 X.laeyis mR
27	1536.5	62.9	1932	9 HOMFLI1A	Y14774 Coturnix co
28	1486.5	60.8	1245	5 AF177538	M93255 Human Flt-1
29	1227	50.2	816	10 S66169	AF177538 Danilo rer
30	1183.5	48.4	762	30 S72621	S66169 Erg-3-immun
31	965	39.5	1431	9 AF327066	S72621 EMS...erg l
32	956.5	39.1	14518	9 AP001423	AF327066 Homo sapi
33	956.5	39.1	100000	9 AP000021	AP001423 Homo sapi
34	956.5	39.1	100000	9 AP000163	AP000021 Homo sapi
35	956.5	39.1	340000	9 AP001731	AP000163 Homo sapi
36	925	37.8	588	9 S72865	AP001731 Homo sapi
37	833.5	34.1	66352	9 HSY17293	S72865 EMS...EMS-e
38	833.5	34.1	187131	2 AP001535	Y17293 Homo sapien
39	833.5	34.1	217382	2 AP001122	AP001535 Homo sapi
40	682	27.9	634	3 SULERG	AP001122 Homo sapi
41	664	27.2	2180	3 AY060316	M1067 Sea urchin
42	581.5	23.8	491	3 HDI31813	AY060316 Drosophila
43	583.5	22.2	482	3 FC0278692	AJ311813 Hediste d
44	520.5	21.3	169741	9 AC097468	AJ278692 Perinereid
45	519.5	21.3	192968	2 AC127107	AC097468 Homo sapi
					AC127107 Rattus no

ALIGNMENTS

RESULT 1

AY05661 LOCUS 1413 bp mRNA linear VRT 17-DEC-2001
DEFINITION Gallus gallus erg isoform C-1-1 mRNA, complete cds; alternatively
spliced.
ACCESSION AY05661
VERSION AY05661.1 GI:17887440
KEYWORDS
SOURCE Gallus gallus.
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 1413)
AUTHORS Iwamoto, M., Higuchi, Y., Koyama, E., Enomoto-Iwamoto, M., Kurisu, K.,
Yeh, H., Abrams, W.R., Rosenblom, J., and Pacifici, M.
TITLE Transcription factor ERG variants and functional diversification of
chondrocytes during limb long bone development
J. Cell Biol. 150 (1), 27-40 (2000)
JOURNAL MEDLINE
PUBMED 20351415
10893256
REFERENCE 2 (bases 1 to 1413)
AUTHORS Iwamoto, M., Higuchi, Y., Enomoto-Iwamoto, M., Kurisu, K.,
Yeh, H., Rosenblom, J., and Pacifici, M.
TITLE The role of ERG (ets related gene) in cartilage development
JOURNAL osteoarthr. Cartil. 9 Suppl A, S41-S47 (2001)
PUBMED 21535378
11680687
REFERENCE 3 (bases 1 to 1413)
AUTHORS Iwamoto, M., Higuchi, Y., Yeh, H., and Pacifici, M.
TITLE Direct Submission
JOURNAL Submitted (05-DEC-2001) Oral Anatomy and Developmental Biology,
Osaka University Faculty of Dentistry, 1-8 Yamadaoka, Suita, Osaka
565-0871, Japan
FEATURES
source Location/Qualifiers
1..1413
/organism="Gallus gallus"
/db_xref="taxon:9031"
37..1392
/note="alternatively spliced; transcription factor"
/codon_start=1
/product="erg isoform C-1-1"
/protein_id="FAL40889.1"
/db_xref="GI:17887441"
/translation="MASTIKELSVSEDSIFECAYGSPHIAKTEMRASSESEGCOT
SKMSPRVQDWLSOPAKAVTIKMCNPNQVGSNSPDSCVAGGKVVSSDVGM
NNGSYMEKHHPPMNTNERVIVADPTLMSTDHVROMLMAVKEYGLPDVILIF
ONIDCKELCKMTKDFORLTPSYNADILSHLYRERGAIFPNTSVYPATORT
TRPDLPYDQARSASWTSHSPTOSKATOPSSYKTEPORQOLDPYOILGPTSSRLA
NPGSGIOLMOLELISDSSNSNCTEGTGGERKMDPDVARMRGKRKPMNY
DKLSRALRYTKNIMTKYHGKRYAKKDFHIGIALQALPHPRESSMYKTPSLPYMS
YHAHPQKMFVAPHPPLCPVTSSFFAAPNPYNSPTGCTYPTRLPAHMPSHLCTY
Y"
BASE COUNT 424 a 370 c 311 g 308 t
ORIGIN
Alignment scores:
Pred. No.: 1,4e-161 Length: 1413
Score: 2444.00 Matches: 451
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0
US-09-902-772-2 (1-451) x AY05661 (1-1413)

Qy 1 MetAlaSerThrIleValGluAlaLeuSerValValSerGluAspGlnSerLeuPheGlu 20
Db 37 ATGGCAAGCACTATTAGAGAGCATTTATCATGGTGGAGAGACAGCTCTGTGTGAG 96
Qy 21 CysAlaTyrGlySerProHisLeuAlaLysThrGluMetThrAlaSerSerSerGlu 40
Db 97 TGTGGCTAGGAGTCGGCCACCTTGCAAGACAGAAATATACAGCCCTCCCTTCACATGAA 156

Qy 41 TyrGlyGlnThrSerIysMetSerProArgValProGlnGlnAspTrpLeuSerGlnPro 60
Db 157 TATGGCAAAACATCAAAAGATGAGACCCGCGCTTCCACAGAGACATGTTATACAGCCC 216
Qy 61 ProAlaArgValThrIleValMetGluCysAspProAsnGlnValAsnGlySerArgAsn 80
Db 217 CCGGCGAGATTACATTAAAGATGAGGTGTAACCAAAACAGGTTAATGGCTAAAGAAAT 276
Qy 81 SerProAspAspCysSerValAlaLysGlyLysMetValSerSerSerAspAsnVal 100
Db 277 TCACGTGATGACAGTCGACGCGGCAAAAGAGGAAATATGGTTATGACGTTCAAGCATGTT 336
Qy 101 GlyMetAsnTyrGlySerTyrMetGluGluLysHisIleProProProAsnMetThr 120
Db 337 GGGATGACATGATGGAAGCATACATGAAAGAAACATATTCGGCTTCAATATGACAAACC 396
Qy 121 AsnGluArgArgValIleValProAlaAspProThrLeuTrpSerThrAspHisValArg 140
Db 397 AATGAACGAAGATTATTGTGCCAGCAGATCTTACGTTATGGACACAGCATGTGACG 456
Qy 141 GlnTrpLeuGluTrpAlaValLysGluTyrGlyLeuProAspValAspIleLeuLeuPhe 160
Db 457 CAGTGGCTGGAGTGGGAGAGGAGATGATGCTTCCAGAGCTGACATCTTGTGTTTC 516
Qy 161 GlnAsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArgLeuThr 180
Db 517 CAGAACATGATGGAAAGAGTGTGTAAATGACCAAAAGATGACTTCCAGAGACTCACG 576
Qy 181 ProSerTyrAsnAlaAspIleLeuLeuSerHisIleuHisTyrLeuArgGluArgGlyAla 200
Db 577 CCGAGCTATACCAATATCTCTCTGTCACACTACACTACTCTCAGAGAGAGAGAGCC 636
Qy 201 ThrPheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgIleThrThrArg 220
Db 637 ACTTTTATTTTCCAAATGATCATGATGTTTACCAGAAAGCAAGCAAGAAATACAAACAAG 696
Qy 221 ProAspLeuProTyrGluGlnAlaArgArgSerAlaTrpThrSerHisSerHisProThr 240
Db 697 CCAGATTACCTTATAGCAGAGAGAGATGACCTGTGACAGATCCACATCCACT 756
Qy 241 GlnSerLysAlaThrGlnProSerSerSerThrValProLysThrGluAspGlnArgPro 260
Db 757 CAGTCAAAAGCTACCAACCATCATCTTTCACACAGTSCCAAAACAGAACCAAGCGCTCT 816
Qy 261 GlnLeuAspProTyrGlnIleLeuGlyProThrSerSerArgLeuAlaAsnProGlySer 280
Db 817 CAGTTAGATCCATTATGAGATTTCTGGACGACAGCAGCAGCGCTTGCAAAATCCAGGAGAT 876
Qy 281 GlyGlnIleGlnLeuTrpGlnPheLeuLeuGluLeuLeuSerAspSerAsnSerAsn 300
Db 877 GGGCAATACACCTATAGGAGAGTTCCTACATGAGCTTCTCGACAGCTCCAAATCCCAAC 936
Qy 301 CysIleThrTrpGluGlyThrAsnGlyLupPheLysMetThrAspProAspGluValAla 320
Db 937 TGCATCACCTGGGAGGAGGCAAAATGGGAGTTCCAGATACAGACCCCTGATATAGTGCT 996
Qy 321 ArgArgTrpGlyGluArgLysSerLysProAsnMetAsnTyrAspLysLeuSerArgAla 340
Db 997 CGGCGTTGGGAGAGAGCAAAACCTAACATGAACTATGACAAACTGACCGCTGCA 1056
Qy 341 LeuArgTyrTyrTyrAspLysAsnIleMetThrLysValHisGlyLysArgTyrAlaTyr 360
Db 1057 CTTCGCTACTACTATGCAAAATATATATATGACTAAAGTTCAAGTCAATGCGCTATGCTAC 1116
Qy 361 LysPheAspPheHisGlyIleAlaGlnAlaLeuGlnProHisProProGluSerMet 380
Db 1117 AATATTGATTTCCAGGAAATCGCTCAGGCCCTCCAGCCCTCCCAAAATCATTCACAG 1176
Qy 381 TyrLysTyrProSerAspLeuProTyrMetSerSerTyrHisAlaHisProGlnLysMet 400
Db 1177 TACAATATACCATACAGACCTCCCTACATGAGATTCCTACATGACACACCCCAAGAGATG 1236
Qy 401 AsnPheValAlaProHisProProAlaLeuProValThrSerSerSerPhePheAlaAla 420

|||||
Db 1237 AACCTTGTAGTCCCATCCCTGCTTCCCGGTAACTCATCCAGCTTTTGTGCTGCC 1296
QY 421 ProAsnProTyrTrpAsnSerProThrGlyIleTyrProAsnThrArgLeuProAla 440
Db 1297 CCTAATCATACGAAATGATCACAACGAGGAGGATCTACCAATACAGGCTGCCAGCT 1356
QY 441 AlaHisMetProSerHisLeuGlyThrTyrTyr 451
Db 1357 GCTCATATGCTTCCCATCTTGCGACCTACTAC 1389
RESULT 2
E31253
LOCUS E31253 1447 bp DNA linear PAT 18-JUN-2001
DEFINITION Protein having cell calcifying inhibitory activity and gene
encoding the same.
ACCESSION E31253
KEYWORDS E31253 1 GI:13025685
SOURCE JP 1999075871-A/1.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 1447)
AUTHORS Hiroyasu, I., Yoshinobu, H., Marjio, P., Joel, R. and Helena, E.
TITLE Protein having cell calcifying inhibitory activity and gene
encoding the same
JOURNAL Patent: JP 1999075871-A 1 23-MAR-1999;
CHUGAI PHARMACEUT CO LTD, HE TRUSTEES OF THE UNIVERSITY OF
PENNSYLVANIA
COMMENT OS Unidentified
PN JP 1999075871-A/1
PD 23-MAR-1999
PR 29-MAY-1998 JP 199816076
PF 18-JUN-1997 US 08/878177, 20-JUN-1997 US 60/050297 PI
HIOYASU IWAMOTO, YOSHINOBU HIGUCHI, MARIJIO PASHIFIKI, PI JOEL
ROZENBUROOM.
PI HELENA E
PC C12N15/09, A61K48/00, C07K14/465, C07K16/18, C12Q1/68//A61K38/00,
PC C12P21/02,
PC C12P21/08, C12N15/00, A61K37/02
CC Strandedness: Single;
CC Topology: Linear;
FH Key Location/Qualifiers
FT source 1. 1447
FT Location/Qualifiers
FT /organism='Unidentified'.
FEATURES
source 1. 1447
BASE COUNT 440 a 374 c 317 g 316 t
ORIGIN
Alignment Scores:
Pred. No.: 1,44e-161 Length: 1447
Score: 2444.00 Matches: 451
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-09-902-772-2 (1-451) x E31253 (1-1447)
QY 1 MetaLaserThrIleLysGluAlaLeuSerValSerGluAspGlnSerLeuPheGlu 20
Db 63 ATGGCAAGCACTATTAAGGAAGCATTTATCAGTGTGAGTGAAGACCAAGCTCTTGTTGAG 122
QY 21 CysAlaTyrGlySerProHisLeuAlaLysThrGluMetThrAlaSerSerSerGlu 40
Db 123 TGTGCTTACGATCGCCCACTTGCAAAAGACGAATGACACCTCTTCCAGTGAA 182
QY 41 TyrGlyGlnThrSerLysMetSerProArgValProGlnGlnAspTrpLeuSerGlnPro 60
Db 183 TATGGCAAAATCAATCAAGATGAGCCGCGGCTTCCCAAGCAGACTGATTATCACAGCCC 242

QY 61 ProAlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsn 80
Db 243 CCGGCCAGAGTATCCATTAAGATGAGTGTAAACCAACAGAGTTAATGGGTCAAGAAAT 302
QY 81 SerProAspAspCysSerValAlaLysGlyLysMetValSerSerSerAspAsnVal 100
Db 303 TCACCTTATGACTGTCACCGGTGGCAAAAGAGGGAATGGTTACAGCTTCAGCAAAAGTT 362
QY 101 GlyMetAsnTyrGlySerTyrMetGluGluLysHisIleProPropoAsnMetThrThr 120
Db 363 GGGATGAACATATGGAACCTATGGAAGAGAGATTTCCGCTCCAAATATATACAAAC 422
QY 121 AsnGluArgArgValIleValProAlaAspProThrLeuTrpSerThrAspHisValArg 140
Db 423 AATGACGAAGAGATTATTTGGCCAGAGATCCATGCTTATGAGACACAGACATGATGCGG 482
QY 141 GlnTrpLeuGluTrpAlaValLysGluTyrGlyLeuProAspValAspIleLeuPhe 160
Db 483 CAGTGGCTGAGTGGGCGAGGGAAGAGATATGCTTCCAGACGTGACATCTTGTGTTCC 542
QY 161 GlnAsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArgLeuThr 180
Db 543 CAGAACATTAATGGAAGAGATTGTGTAATGACCAAGATGACTTCCAGAGACTCAGC 602
QY 181 ProSerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGluArgGlyAla 200
Db 603 CCGAGTATGACGAGATATCCCTGCTGACACCTACATACCTCAGAGAGAGAGAGGCC 662
QY 201 ThrPheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgIleThrThrArg 220
Db 663 ACTTTATTTTCCAAATATCATCATGTTTACCACCAACAGCAAGCAAAATTAACAAGG 722
QY 221 ProAspLeuProTyrGluGlnAlaArgArgSerAlaTrpThrSerHisSerHisProThr 240
Db 723 CCAGATTACTTATGAGCAAGCGAGAGATCAGCGTGGAGATCTACAGCATCCCACT 782
QY 241 GlnSerLysAlaThrGlnProSerSerSerThrValProLysThrGluAspGlnArgPro 260
Db 783 CAGTCAAAAGCTACCCCAACCATCATCTTCAACAGTGGCCCAAAAGACAGACAGCTGCT 842
QY 261 GlnLeuAspProTyrGlnIleLeuGlyProThrSerSerArgLeuAlaAsnProGlySer 280
Db 843 CAGTTTATATCTTATCATCATGATCTTGGACCGACGCGCTTGGCAAAATCCAGGAGGT 902
QY 281 GlyGlnIleGlnLeuTrpGlnPheLeuLeuGlnLeuLeuSerAspSerAsnSerAsn 300
Db 903 GGGCAGATACAGCTATGAGCAGTTCCTACTGAGGCTTGTGCGGACAGCTCCAACTCCAC 962
QY 301 CysIleThrTrpGluGlyThrAsnGlyLysPheLysMetThrAspProAspGluValAla 320
Db 963 TGCATCACCTGGGAGCGGCACAANTGGGAGTTCAAGATGACAGACCTCGATGAAGTGGCT 1022
QY 321 ArgArgTrpGlyGluArgLysSerLysProAsnMetAsnTyrAspLysLeuSerArgAla 340
Db 1023 CGGCGTGGGGAAGAGGAAGCAAAACCTTAACATGATGCAAAACCTCAGCCGCGCA 1082
QY 341 LeuArgTyrTyrTrpAspLysAsnIleMetThrLysValHisGlyLysArgTyrAlaTyr 360
Db 1083 CTTCGCTACTACTATGCAAAAATATTATATGACTTAAGTTATGATTAGCTTAAGCTATGCTTAC 1142
QY 361 LysPheAspPheHisGlyIleAlaGlnAlaLeuGlnProHisProGlnIleSerSerMet 380
Db 1143 AAATTGATTTCCACGGAATCGCTCAGGCGCTCAGCGCTCACCTCCAGATATCATCATG 1202
QY 381 TyrLysTyrProSerAspLeuProTyrMetSerSerTyrHisAlaHisProGlnLysMet 400
Db 1203 TACAAAATACCATCAGACCTCCCTTACATGATGATTCATCATGACACCCCGAAGATG 1262
QY 401 AsnPheValAlaProHisProProAlaLeuProValThrSerSerSerPhePheAlaAla 420
Db 1263 AACCTTTGAGTCCCATCCCTGCTTGTGCGGTAACTCATATCCAGCTTTTGTGCTCC 1322

OY	421	ProAenProtTyTrTPAnSerPerThrcrGlyGIyIeTYPrProAsnThrArgLeuProAla	440
Dd	1323	CCTATCCACTAGCGAAATTACCAACAACGTGGAGCATCTACCACCAATAACGAGTGCACGCT	1382
OY	441	AlahIsMetPSeHsrHisreugLythrTYrrYr	451
Db	1383	GCTCATTGCCCTTCCTCCATCTTGCCACCTACTAAC	1415
RESULT 3			
LOCUS	GGERG	1516 bp	mRNA linear VRT 27-APR-1995
DEFINITION	G.gallus ERG mRNA.		
ACCESSION	X77159		
VERSION	X77159.1 GI:790439		
KEYWORDS	ERG gene.		
SOURCE	chicken.		
ORGANISM	Gallus gallus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauaria; Aves; Neognathae; Galliformes; Phasianidae;		
AUTHORS	Phasianidae; Gallus. Dhoradain,P., Dewitte,F., Desbiens,X., Stehelin,D. and Mesodermal expression of the chicken erg gene associated with precartilaginous condensation and cartilage differentiation Mech. Dev. 50 (1), 17-28 (1995)		
TITLE	Mesodermal expression of the chicken erg gene associated with precartilaginous condensation and cartilage differentiation		
JOURNAL	Mech. Dev. 50 (1), 17-28 (1995)		
MEDLINE	95329425		
PUBMED	7605748		
REFERENCE	2 (bases 1 to 1516) Dutergue-Coquillaud,M. Submitted (17-JAN-1994) M. Dutergue-Coquillaud, CNRS UA 1160, Oncologie Moléculaire, Institut Pasteur, 1 rue Calmette, 59019 Lille, FRANCE		
AUTHORS			
TITLE			
JOURNAL			
FEATURES			
SOURCE	Location/Qualifiers 1..1516 /organism="Gallus gallus" /db_xref="taxon:9031" /tissue_type="spleen" /clone_lib="lambda gt10" /dev_stage="adult" 63..1499 /gene="ERG" 63..1499 /gene="ERG" /codon_start=1 /protein_id="CAA54404.1" /db_xref="GI:790440" /db_xref="SWISS-PROT:Q90837" /translation="MASTIKKALSVESEDLSFEACAYSGPHIAKTENTASSSEFYGOT SKMSPREPOODMLSQPPARVLIKMECHNDNGSRNSDDCSVAKGKMSSSDNVGM NYGYMEERKHIPPENMTNERRVLPADPTLMSDHYROMLEAVKEGLPDVDILF ONIDGKETLKWTKDDFORITERSYNADILLSHLHTLDHEPLPLTLTDVDALONSPLU MHANTGATPIFPNTSYYPATORTTRPRPLEQARRSMWTSHIPTOKATOPESS STVRTEQRQLDPYOILIGFTSSRLNPNPGGQIQLOWFLELLSDSSNSKCITWEGETE NGEKMRDDPEVARMGERRKSRRPMNKDRSLRALRYYYDNKNIMTVGKRAYSFFDFHF GIALQHPHPRESSMYXPSDLPYMNSYHAHPQKMNFVAPHRPALPVTSSSFAPAAMPB YMNSTGITYPNTFLPAAHMPSSHLCITY"		
BASE COUNT	458 a 392 c 327 g 339 t		
ORIGIN			
Alignment Scores:			
Pred. NO.:	1.73e-159	Length:	1516
Score:	2414.50	Matches:	450
Percent Similarity:	94.14%	Conservative:	0
Best Local Similarity:	94.14%	Mismatches:	1
Query Match:	98.79%	Indels:	27
DB:	5 Gaps:		1
DS-09-902-772-2 (1-451) x GGERG (1-1516)			
OY	1	MetaIaserTHrIIIdysGlUAleUservAIvAlserGUASgClSerleudheGu	20

Db	63	ATGGCAAGCACTATTAAAGGAAGCATTTTCAGTGTGTGATGAGACCAAGCTCTTTGTTGAG	122
Oy	21	CysAlaTyrGlySerProHisIleuAlaLysThrGluMetThrAlaSerSerSerGlu	40
Db	123	TTGTGGCTAAGGATGGCCCACTTGGCAAGACAGAAATGACAGCCCTCTTCAGTGAA	182
Oy	41	TyrGlyGlnThrSerLysMetSerProArgValProGlnIAspTrpLeuSerGlnPro	60
Db	183	TTATGGGCAAAACATCAAAAGATAGCCGCCGGCTTCCACAGAGACAGCTTATCACAGCCC	242
Oy	61	ProAlaArgValThrIleLysMetGluCysAsnProAsnGlnAlaAsnIleSerArgAsn	80
Db	243	CCGGCCACAGTTTACCATTTAGATGAGTGTACCCCAACAGGTATATGGGTCAAGGAAT	302
Oy	81	SerProAspAspCysSerValAlaLysGlyGlyLysMetValSerSerSerAspAsnAl	100
Db	303	TTACCTGTATGATCGTAGGGTGGCAAAAGAGGAGAAATGGTTAGCAGTTCCACAAATGTT	362
Oy	101	GlyMetAsnTyrGlySerTyrMetGluGluLysHisIleProProAsnMetThrThr	120
Db	363	GGGATGAACCTATGGAAGCTACATGGAAGAAACAAATATCCCGCTCCAAATGTGACAAAC	422
Oy	121	AsnGluArgArgValIleValProAlaAspProThrLeuTrpSerThrAspHisValArg	140
Db	423	AATGACACAAAGATTATTGTGCCAGCAAGATCTTACGTATTGGAGCACAAACCATTTACGG	482
Oy	141	GlnTrpLeuGlnTrpAlaValLysGluTyrGlyLeuProAspValAspIleLeuLeuPhe	160
Db	483	CAGTGGCTGGAGTGGCGCAGTAAAGAGATAGGGCTTCCAGACGTGACATCTTGTGTTTC	542
Oy	161	GlnAsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArgLeuThr	180
Db	543	CAGAACATTTGATGGGAAAGATGTTGTGTAATAATGACCAAGATGACTTCCAGAGACTCAG	602
Oy	181	ProSerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGlnArg-----	198
Db	603	CCGACCTATTAACGGAAGATATCTTCGTGTCCACCTTACACTACCTCAGAGAGACTCTT	662
Oy	198	-----	198
Db	663	CCACATTTGACTTCAGATGATGATGTTGATGAAGGCCCTTACAAAACCTCCACGGTTAATGCAT	722
Oy	199	-----GlyAlaThrPheIlePheProAsnThrSerValTyrProGlnAla	213
Db	723	GCTAAGAAACACAGAGAGAGCCACTTTTATTTTCCAAAATACATCAGTTTACCACAAACGA	782
Oy	214	ThrGlnArgIleThrThrArgProAspLeuProTyrGluGlnAlaArgArgSerAlaTrp	233
Db	783	ACGCAAAATAATTAACAAAGGCCAGATTATCCATTATAGCACAAGCGAGGAGTACACGGG	842
Oy	234	ThrSerHisSerHisProThrGlnSerLysAlaThrGlnProSerSerSerThrValPro	253
Db	843	ACGATCTCACACCCATCCCACTCAGTCAAAAGCTAACCCAAACATCTTCAACAGATGCC	902
Oy	254	LysThrGluAspGlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSerSer	273
Db	903	AAAACAGAAAGACAGCGCTCCATGATGATCTTATCAGATTCCTTGAGCCGACGACGAGC	962
Oy	274	ArgLeuAlaAsnProGlySerGlyGlnIleGlnLeuTrpGlnPheLeuLeuGluLeuLeu	293
Db	963	GCTCTTGGAAATCCAGGAGAGGGGCAATACACCTATGCACTTCTATCGAGGCTTCTG	1022
Oy	294	SerAspSerSerAsnSerAsnCysIleThrTrpGluGlyThrAsnGlyValPheLysMet	313
Db	1023	TTGGACACTCTCACTCCAACTGATCATTACCTGGAGGCGACAAATGGGGAGTTCAAGATG	1082
Oy	314	ThrAspProAspGluValAlaArgArgTrpGlyGluArgLysSerLysProAsnMetAsn	333
Db	1083	ACAGACCTGTATGAAGTGGCGTGGGGAGAGAGAGAAAGCAAAACCTAACATGAAAC	1142
Oy	334	TyrAspLysLeuSerArgAlaLeuArgTyrTyrTyrAspLysAsnIleMetThrLysVal	353
Db	1143	TATGCAAAATCACCGGTGCACTTGGTACTACTATGACAAAAAATATATGTCTAAATCTT	1202

QY	354	HISGLYLSVATGYTALATYRILYSHPHSHSGIYLIALAGINALALEUINPRO	373		
Db	1203	CATGGTAAACGGTAAAGCCCTACAAATTTATTTCCACGGAAATGCTCAGGCCCTCAGGCT	1262		
QY	374	HISPROGILSERSERMETYLRSYTRPROSERASPLEPROTYRMESESRITYR	393		
Db	1263	CACCTTCAGATTCATTCATGTACAAATACCATAGACCTCCCTACATGAGTTCTTAC	1322		
QY	394	HISALAHISPROGILNLYSMETASNPNHEVALALAPROHISPROPROALAUEROVALTHR	413		
Db	1323	CATGACACACCCCAAGATGACACTTTGTAGTCTCCCAATCCCGCTTGCCCGGTAAACC	1382		
QY	414	SERSESRPHEPHEPHEALALAPROASNPROTYRTPASUSERPROTHGLYGLYILETYR	433		
Db	1383	TCATCCACCTTTTGTCTGCCCTCATCTCATCTGGAATTCCACAACTGGAAGCATCTAC	1442		
QY	434	PROASNTPRARGLEUPROALAHISMETPROSEGRHSLSENGLYTHRTYR	451		
Db	1443	CCCAATACCAGGCTGCCACCTGCTCATATGCTTCCCATCTTGGCAGCACTACTAC	1496		
RESULT 4					
E31254		1528 bp	DNA	linear	PAT 18-JUN-2001
LOCUS					
DEFINITION	E31254	Protein having cell calcifying inhibitory activity and gene			
ACCESSION	E31254	encoding the same.			
VERSION	E31254.1	GI:13025686			
KEYWORDS	JP 1999075871-A/2.				
SOURCE	unidentified.				
ORGANISM	unidentified.				
REFERENCE	1 (bases 1 to 1528)				
AUTHORS	Hltroysau,I., Yoshinobu,H., Marjio,P., Joel,R. and Helena,E.				
TITLE	Protein having cell calcifying inhibitory activity and gene encoding the same				
JOURNAL	Patent: JP 1999075871-A 2 23-MAR-1999;				
	CHUGAI PHARMACEUT CO LTD, HE TRUSTEES OF THE UNIVERSITY OF PENNSYLVANIA				
COMMENT	OS Unidentified				
	PN JP 1999075871-A/2				
	PD 23-MAR-1999				
	PF 29-MAY-1998 JP 1998166076				
	PR 18-JUN-1997 US 08/878177, 20-JUN-1997 US 60/050297 PI				
	HITROYASU IWAKOTO, YOSHINOBU HICUCHI, MARIITO PASHIYIKI, PI JOEL ROZENUDROM,				
	PI HELENA E				
	PC C12N15/09,A61K48/00,C07K14/465,C07K16/18,C12Q1/68//A61K38/00,				
	PC C12P21/02,				
	PC C12P21/08,C12N15/00,A61K37/02				
	CC Strandedness: Single;				
	CC Topology: Linear;				
	EH Key				
	FT source				
	FT				
FEATURES					
source					
	location/Qualifiers				
	1..1528				
	/organism='unidentified'				
	/db_xref='taxon:32644'				
BASE COUNT	464 a	395 c	330 g	339 t	
ORIGIN					
Alignment Scores:					
Pred. No.:	1,75e-159	Length:	1528		
Score:	2414.50	Matches:	450		
Percent Similarity:	94.14%	Conservative:	0		
Best Local Similarity:	94.14%	Mismatches:	1		
Query Match:	98.79%	Indels:	27		
DB:	6	Gaps:	1		
US-09-902-772-2 (1-451) x E31254 (1-1528)					
QY	1	MetalaSerThrIleLysGluAlaLeuSerValValSerGluAspGlnSerLeuPheGlu	20		

Db	63	ATGCCAAGCCTATTAAAGGAACATTATCATCGTGTGAGTGAAGACCACTGCTTGTTCAG	122
Qy	21	CysAlaItyrGlySerProHisLeuAlaIysThrGluMetThrAlaSerSerSerGlu	40
Db	123	TGTGCTACAGGATGCGCCCACTTTCGAAGAGACGAATAATACAGCCTCTCTTCCAGTGA	182
Qy	41	TyrGlyGlnThrSerLysMetSerProArgValProGlnInAspTrpLeuSerGlnPro	60
Db	183	TATGGCAAAACATCAAAAGATGAGACCGCGCGGTTCACCACAGAGCTGTATACAGACC	242
Qy	61	ProAlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsn	80
Db	243	CCGGCCAGAGTTACACTTAAGATGAGGTGTAAACCCAAACAGGTTAATGGTCAAGAAAT	302
Qy	81	SerProAspAspCysSerValAlaIysGlyLysMetValSerSerSerAspAsnVal	100
Db	303	TCACCTGATGATCTACACCGCTGGCCAAAAGAGGAAAAATGTTATGCAAGTTCCACACATGTT	362
Qy	101	GlyMetAsnTyrGlySerTyrMetGluGluLysHisIleProProAsnMetThrThr	120
Db	363	GGGATGAACATATGGAACCTACATGGAAGAAGACATATTCGCGCTTCAATATGACAAAC	422
Qy	121	AsnGluArgArgValIleValProAlaAspProThrLeuTrpSerThrAspHisValArg	140
Db	423	AATACAGCAAGATTAATGTGGCCAGACAGATCCATGTTATGGAGACACACACATGTACGG	482
Qy	141	GlnTrpLeuGluTrpAlaValLysGluTyrGlyLeuProAspValAspIleLeuLeuPhe	160
Db	483	CAGTGGCTCGAGTGGGACGAGAAAGAGATATGCTTCCAGACGTGGACATCTTGTGTTTC	542
Qy	161	GlnAsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArgLeuThr	180
Db	543	CAGAACTATGATGGAAAGAGTTGTGTAAATGACCAAAAGATGACTTCCAGAGACTCAG	602
Qy	181	ProSerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGluArg-----	198
Db	603	CCGAGCTATTAACGACGATATCTCTCGTGCACACCTACCTACCTCAGAGACCTCCCTT	662
Qy	198	-----	198
Db	663	CCACATTTGACTTCATGATGATTTAAAGCCCTTAAACAACTCTCCACGGTTAATGCAT	722
Qy	199	-----GlyAlaThrPheIlePheProAsnThrSerValTyrProGluAla	213
Db	723	GCTAGAAACACAGAGAGAGCCACTTTATTTTCCAAATACATCAGTTTACCCAGAAACA	782
Qy	214	ThrGlnArgIleThrThrArgProAspLeuProTyrGluGlnAlaArgArgSerAlaTrp	233
Db	783	ACGCAAAAGAAATAAACAACAAGCCAGATTTAACTTATGACCAAGCAGAGATCAGCGTG	842
Qy	234	ThrSerHisSerHisProThrGlnSerLysValAlaThrGlnProSerSerSerThrValPro	253
Db	843	ACGAGTCACAGCCCTCCACTCTCAGTCAAAAGCTACCCCAACCATCTTCAACACTGGCC	902
Qy	254	LysThrGluAspGlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSerSer	273
Db	903	AAAAACAGAGCCAGCGCTCCTCAGTTATGATCCTTATCAGATTCCTTGGACCCAGCAGC	962
Qy	274	ArgLeuAlaAsnProGlySerGlyGlnIleGlnLeuTrpGlnPheLeuLeuGlnLeuLeu	293
Db	963	CGTCTGCAATTCAGGAGGAGTGGGACAGATTAACACTTATGCGACTTCTACTGGAGCTCTG	1022
Qy	294	SerAspSerSerAsnSerAsnCysIleThrTrpGluGlyThrAsnGlyGluPheLysMet	313
Db	1023	TTCGACAGCTCCAACTCCAACTCATCTACCTCTGGAGGCGACAAAAGGGGAGTTCACAGATG	1082
Qy	314	ThrAspProAspGluValAlaArgArgTrpGlyGluArgLysSerLysProAsnMetAsn	333
Db	1083	ACAAGCCTGATGAAGTGGCTCGGCGTTGGGAGAGAGAAAAGCAAACTTACATGTGAC	1142
Qy	334	TyrAspLysLeuSerArgAlaLeuArgTyrTyrTyrAspLysAsnIleMetThrLysVal	353

Db 1143 TATGACAACTCAGCCGTCACCTTCCTACTACTATGACAAAAATATATGACTAAAGTT 1202

QY 354 HHSGLYLSARGTYTALAATYLYSPHEASPHENHSGLYTLAAGLALALEUINPRO 373

Db 1203 CATGGTAAACGGTATCTCCACAAATTTGATTTCCACGGAATCCCTCAGGCCCTCCAGCCCT 1262

QY 374 HHSProGlySerSerMetTYLYSTYRProSerAspLeuProIyrmetserserTYR 393

Db 1263 CACCCCTCCAGATCATCTGACAAATACCCCTCAGACCTCCCTCCATCAGAGTTCCTTAC 1322

QY 394 HHSALAHSProGlyLysMetAspPheValAlaProHisProProAlaLeuProValThr 413

Db 1323 CATGCACACCCCGAAGATGAACCTTGTAGCTCCCAATCCCTCTTGGCCCGTAACCC 1382

QY 414 SerSerSerPhePheAlaAlaProAsnProTYRTTPASerProThrGlyIleTYR 433

Db 1383 TCATCAGCTTTTTCCTCCCTCCATTCATCTGGAATTCACCACTGAGCAGCATCTTAC 1442

QY 434 ProAsnThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTYR 451

Db 1443 CCCAATACAGGCTGCCAGCTGCTCATATGCTTCCCATCTTGCCACCTACTAC 1496

RESULT 5

AB073080 2209 bp mRNA linear ROD 17-OCT-2001

LOCUS AB073080

DEFINITION Mus musculus Erg mRNA, mouse homolog of Human ets-related gene ERG, transcript variant 3, complete cds.

ACCESSION AB073080

VERSION AB073080.1 GI:16197544

KEYWORDS Mus musculus cDNA to mRNA, clone:Erg-3.

SOURCE Mus musculus

ORGANISM Mus musculus

REFERENCE 1 Ozawa,R., Noguchi,H., Taylor,T.D., Takeda,T., Hattori,M. and Sakaki,Y.

AUTHORS Mus musculus Erg mRNA

TITLE Published Only in database (2001)

JOURNAL 2 (bases 1 to 2209)

AUTHORS Ozawa,R., Noguchi,H., Taylor,T.D., Takeda,T., Hattori,M. and Sakaki,Y.

TITLE Direct Submission

JOURNAL Submitted (16-OCT-2001) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-Chou,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:hattori@gs.c.riken.go.jp, URL:http://hgp.gs.c.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)

AUTHORS

FEATURES

SOURCE Location/Qualifiers

1..2209

/organism="Mus musculus"

/db_xref="taxon:10090"

/chromosome="16"

/clone="Erg-3"

1..2209

/gene="Erg"

148..1608

/gene="Erg"

/note="homolog of human ets-related gene (ERG), transcript variant 3"

/codon_start=1

/protein_id="BAB69950.1"

/db_xref="GI:16197545"

/translation="MIDVPPDAHIKEALSVSHDSLFECAVGTPIHAKTEMTASSSSDYGOTKMSRPVPODMLSOAPARVITIKECNSOYGNRSNPDSECVKGGKMGSPDYVGMSTGYMEKHVPPNMTNERRIVPADPTLMSRDHYKMLKMAVKEVGLLDVYLILRQNTGKEKELCKTKDDFORLTPSYNADILSLHLIRETPLEHLISDDYDKALQNSPRLHAKNTGGAATIFPNSTVYEPATQRTTRPDLPEPPRSAMTGHSHLTPQSKAOPSPSAVKTEDORADPYQIILPTSSRLANPSGQIOLMOLFLELSDSSNSNCTIWEETNGEFTKMTDPDEVARWGERSKSNMNDKLSRLARYYDINIMTKYGRYAKFEHGIQALQHPPESSLYKYPEDLPMGYSVHAHQPKMNFVSPHPALPTTSSSFAFSPNPMYSNGGTYPNRLPASHHPSHLGTY"

BASE COUNT 606 a 608 c 545 g 450 t

ORIGIN

Alignment Scores:

Pred. No.: 5,18e-149

Score: 2267.00

Percent Similarity: 90.59%

Best Local Similarity: 88.08%

Query Match: 92.768

DB: 10

Gaps: 2

US-09-902-772-2 (1-451) x AB073080 (1-2209)

QY 2 AAlaSerThrIleLysGluAlaLeuSerValValSerGluAspGlnSerLeuPheGluCys 21

Db 172 GCAGCTCATATTAAAGAGCGCTTGTCAGTTGTGAGGAGAGACAGACAGTATTGTAGTGT 231

QY 22 AAlaTYGlySerProHisLeuAlaLysThrGluMetThrAlaSerSerSerGlyTYR 41

Db 232 GCCTACGGAAAGCCACACCTGGCTAAGACAGAGATACCCGATCTCTTCCAGTACTAT 291

QY 42 GlyGlnThrSerLysMetSerProArgValProGlnGlnAspThrLeuSerGlnProPro 61

Db 292 GGCACAGACATCCAAAGATGAGTCCAGAGTCCCTCAGACAGAGCTGGCTGTCAAGCCCA 351

QY 62 AAlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsnSer 81

Db 352 GCCAGGCTCACCATVCAAGATGAGTCCAAACCCATGACAGTGTGATGCTCCAGGAACCTCA 411

QY 82 ProAspAspCysSerValAlaLysGlyGlyLysMetValSerSerSerAspAsnValGly 101

Db 412 CTGTAGTAGAGTGCAGTGTGAACAAAGGTGGAGAGATGTGGAGCCCGGATCTGTGGGG 471

QY 102 MetAsnTYGlySerTYrMetGluGluLysHisIleProProAsnMetThrThrAsn 121

Db 472 ATGAGCTACGGACGATACATATGAGAGAGAAAGCATGTGCCCTCCCAATGTGACCAAAAT 531

QY 122 GluArgArgValIleValProAlaAspProThrLeuTrpSerThrAspHisValArgGln 141

Db 532 GAGCGGAGAGTATCTCTCCCTGAGATCTTCTGTGGAGCAGAACCATGTCCGACAG 591

QY 142 TrpLeuGluTrpAlaValAlaLysGluTYRGLysLeuProAspValAspIleLeuLeuPheGln 161

Db 592 TGGCTGGAATGGCGGTGAAAGATATGGCTCTCTCGATGTGAGCTCTTACTATTTCAG 651

QY 162 AsnIleAspGlyLysGluLeuLysLysMetThrLysAspAspPheGlnArgLeuThrPro 181

Db 652 AATATGATGAGGAAGAGAGCTGTGCAGAAAGATGACAAAGATGACATCTCCAGGGCTCACGCCG 711

QY 182 SerTYrAsnAlaAspIleLeuLeuSerHisLeuHisTYrLeuArgGluArg 198

Db 712 ACCTCAATGCGCAGCATCTTCTCTCACATCTCCACTACCTAGAGAGACTCCCTTCCA 771

QY 198 ----- 198

Db 772 CATCTGACTTCCGATGACCTTGATTAAGCTTTACAAAACCTCTCCACGGTTAATGCATGCC 831

QY 199 -----GlyAlaThrPheIlePheProAsnThrSerValTYRProGluAlaThr 214

Db 832 AGAAACACAGGGGGTGCACCTTTATTTTCCCAATACTCTAGTATATCCCAAGCTACG 891

QY 215 GlnArgIleThrThrArgProAspLeuProTYRGLuGlnAlaArgArgSerAlaTrpThr 234

Db 892 CAAAGATTTACAACTAGCCACAGATTTCCTTATGAGCCCTCCAGAGATGACCTCGAGACC 951

QY 235 SerHisSerHisProThr---GlnSerLysAlaThrGlnProSerSerSerThrValPro 253

Db 952 GGCACAGCCACCTCAACCCCTCAGTCCAAAGCTGCTCAGCCATCTCCCTCAGCTGCC 1011

QY 254 LysThrGluAspGlnArgProGlnLeuAspProTYRGLuGlnIleLeuGlyProThrSerSer 273

Db 1012 AAACCTGAAGACACAGCTCTCTAGTATGATCTTACACAGATCTCTGGAGCCAGCACTAC 1071

QY 274 ArgLeuAlaAsnProGlySerGlyGlnIleGlnLeuTrpGlnPheLeuLeuLeuLeu 293

Db 1072 CGCCTTCCTAATCCAGTAGTGGCCAGATCCAGTCTGGCAGTTCCTCCGACCTCCTG 1131
 QY SeraspserAenSerAenCys11eThrTrpGluGlyThrAenGlyGluPheLysMet 313
 Db 1132 TCAGACAGCTCCAACTCACTGCATCCTCGGAGAACACACGGGAGTCAACATG 1191
 QY 314 ThrAspProAspGluValAlaArgArgTrpGlyGluArgLysSerLysProAsnMetAsn 333
 Db 1192 ACAGACCAGGACGAGTGGCTGGCGCTGGGGGAGAGAACACCAAGCCACATGTAAC 1251
 QY 334 TyrAspLysLeuSerArgAlaLeuArgTrpTyrTyrAspLysAsnIleMetThrLysVal 353
 Db 1252 TATGACAGAGCTCAGCCGCGCTCCGCTACTACTACGACAAACATCATGACCAAGGTG 1311
 QY 354 HisGlyLysArgTrpAlaTrpLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnPro 373
 Db 1312 CAGCGGAGAGGCTACGGCTTCAAGTTGACTCCACGGGAGTGGCCAGGCCCTCCAGACC 1371
 QY 374 HisProGluSerSerMetTyrLysTyrProSerAspLeuProTyrMetSerSerTyr 393
 Db 1372 CACCTCTCTAGTCTCCCTGACAAAGTACCCCTCCGACCTGCGCATGATGGCTCTAT 1431
 QY 394 HisAlaHisProGlnLysMetAsnPheValAlaProHisProValAlaLeuProValThr 413
 Db 1432 CAGCGCCACCCCGAGAGAGTAACTTGTGTCTCCACCCCTCCGCTCCACAGTACA 1491
 QY 414 SerSerSerPhePheAlaAlaProAsnProTyrTrpAsnSerProThrGlyIleLys 433
 Db 1492 TCTTCCAGTTCTTGTCTCCCGAACCCATACGAAATTCACCGCATGGGGGAGTCTAC 1551
 QY 434 ProAsnThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyr 451
 Db 1552 CCGAACACTAGGCTCCCGACGACCATATGCTCTCACCCTGGGACCTACTAC 1605
 RESULT 6
 XLAJ4126 2012 bp mRNA linear VRT 21-JAN-2000
 LOCUS XLAJ4126 xenopus laevis erg gene (erg_E).
 DEFINITION AJ224126
 ACCESSION AJ224126.1 GI:5420047
 VERSION 1
 KEYWORDS ERG gene; transcription factor.
 SOURCE Xenopus laevis.
 ORGANISM Xenopus laevis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 Xenopodinae; Xenopus.
 1 (bases 1 to 2012)
 Balczinger M., Mager-Heckel A.M. and Remy P.
 XI erg: expression pattern and overexpression during development
 Dev. Dyn. 216 (4-5), 420-433 (1999)
 JOURNAL MEDLINE 20096678
 PUBMED 10633861
 REFERENCE 2 (bases 1 to 2012)
 AUTHORS Balczinger M.
 TITLE Direct Submission
 JOURNAL Submitted (24-FEB-1998) Balczinger M., UP9005, MMDCD, Centre
 National de la Recherche Scientifique, 15, RUE Rene Descartes,
 67084, FRANCE
 FEATURES
 source
 1. .2012
 /organism="Xenopus laevis"
 /db_xref="taxon:8355"
 /clone="lambda ZAPri-erg E"
 /dev_stage="stage 30 embryo"
 /gene="erg"
 /gene="erg"
 /gene="erg"
 /function="transcription factor"
 /codon_start=1
 CDS
 gene
 CDS

/protein_id="CAB46567.1"
 /db_xref="GI:5420048"
 /db_xref="SPTREMBL:O9M629"
 /translation="MIOGTWDPSPAHKIREALSVSEDSOLFECTYGRPHLTKTEMTASS
 SDDYGTGSKSPRPVROODWLSQPPSRRTIMECNPSVNSPPDCISGKSKMAG
 GSDNNGNNGSYIEEKHISPNMNTNRRTVPRADPLWSTDVROWLEWAIETGLP
 DVVLLQNIDGELCKMKTEDEHRLTPSNADILSLHLYLERGASFTLPNSVYQ
 DANORIPESRODLSYEPSRSAMNHPAPSKASQSTYVPTKEDPRQDLPYQILGPT
 SSRLANSGQIOLMOFLLELLSDSSNSNITWEGTGEFKMTDPDEVARMGERSK
 PNNYDKLSRLRYDYDKNMTYHGKRYRYKDFEHLIAQALQPHPESTMRYKPSFL
 PYMSVYHAHPQKNFVAPHPALPYVSSSFPAFPMWNSPTGSIYENTLPSHMS
 HLGTTY"
 BASE COUNT 634 a 479 c 395 g 504 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 7 84e-144 Length: 2012
 Score: 2192.00 Matches: 402
 Percent Similarity: 94.63% Conservative: 21
 Best Local Similarity: 89.93% Mismatches: 22
 Query Match: 89.69% Indels: 2
 DB: 5 gaps: 2
 US-09-902-772-2 (1-451) x XLAJ4126 (1-2012)
 QY 5 IleLysGluAlaLeuSerValAlaSerGluAspGlnSerLeuPheGluCysAlaTyrGly 24
 Db 258 ATCAAGAGCTCTTCAGTGTGAGTGAAGATCAGTGTATTCGATGACCTAGCGA 317
 QY 25 SerProHisLeuAlaLysThrGluMetThrAlaSerSerSerSerGluTyrGlyGlnThr 44
 Db 318 ACGCCACATCTTACTAAACGAGATGACCCCATCTTCCTCAGTGCATAGGCGCAAC 377
 QY 45 SerLysMetSerProArgValProGlnGlnAspTrpLeuSerGlnProValArgVal 64
 Db 378 TCGAAATGAGTCCAGCGTGGCTCAACACAGCTGGCTCTCAGCCCTCCATCAGGCTG 437
 QY 65 ThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsnSerProAspAsp 84
 Db 438 ACCATCAAGATGGAATGCAACCCCTAGTCAAGTCAATGATCAAGAGCTCCCGGATGAC 497
 QY 85 CysSerValAlaLysGlyGlyLysMetValSerSerSerAspAsnValGlyMetAsnTyr 104
 Db 498 TGCAGTATAGCAAAAGAGCAAGCAAGTGGTGGAGCGGTGATGTTGGTGAATGAATAT 557
 QY 105 GlySerTyrMetGluGluLysHisIleProProProAsnMetThrAsnGluArg 124
 Db 558 GGCAGCTACATAGCAAGAAACACATTTCTCCCAACATGACACCAATGAGAGGAGA 617
 QY 125 ValIleValProAlaAspProThrLeuTrpSerThrAspHisValArgGlnTrpLeuGlu 144
 Db 618 GTGATTTACTGCTGACCTACTTGTGGAGCAGATCATGTCAGACAGTGGCTGAG 677
 QY 145 TrpAlaValLysGluTyrGlyLeuProAspValAspIleLeuLeuPheGlnAsnIleAsp 164
 Db 678 TGGCAATTAAGAAAGATACGGCTTCCAGATGTCATGTGCTCCGTTCCAAACATGAC 737
 QY 165 GlyLysGluLeuCysLysMetThrLysAspAspPheGlnArgLeuThrProSerTyrAsn 184
 Db 738 GGCAGGAATTTAGCAAGATGACCAAGAAAGATTTCCATGACATGACGCCCAAGTACAT 797
 QY 185 AlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGluArgGlyAlaThrPheIlePhe 204
 Db 798 GCTATATCTCTCTCGCATCTTCATCTACCTCGAAGAAAGAGTGGCTTTATTTTC 857
 QY 205 ProAsnThrSerValTyrProGluAlaThrGlnArgIleThrThrArgProAspLeuPro 224
 Db 858 CCAAACTCATCCGTTTATCAAGACGCAACCAAGAAATCCCAACAGCAAGATTTATCA 917
 QY 225 TyrGluGlnAlaArgArgSerAlaTrpThrSerHisSerHisProGlnGlnSerLysAla 244
 Db 918 TATGAGCCCTCAAGAGATGCTGTTGACAAATCATCTCTGCACCA---CCATCAAAAGCA 974

QY 245 ThrGlnProSerSerSerThrValProLysThrGlnuAspGlnArgProGlnuLeuAspPro 264
Db 975 TCTCAACCA---TCTAACCAAGTCCCAAAACAGAACCCACGCGCACAGCTAGATCCT 1031
QY 265 TTTGlnIleuGlyProThrSerSerArgLeuAlaAsnProGlySerGlyGlnIleGln 284
Db 1032 TATCAGATCTTGGGCAACAGACGCGACTTGCAATCCAGGAAGTGGGAGATTGAG 1091
QY 285 LeuTyrGlnPheLeuLeuGlnuLeuLeuSerAspSerSerAsnSerAsnCysIleThrTrp 304
Db 1092 CTTGGCAATCTTACTGCAAGCTGCTTGGATGATGTCACCACTCAATGATTAATCTGG 1151
QY 305 GlnuGlyThrAsnGlyGlnuPheLeuMetThrAspProAspGlnuValAlaArgArgTrpGly 324
Db 1152 GAAGGAACCAATGGAGAAATTAAGATGACCGATCTGATGAAGTGGCAAGCGTTGGGG 1211
QY 325 GlnuArgSerSerLysProAsnMetAsnTyrAspLysLeuSerArgAlaLeuArgTyrTrp 344
Db 1212 GAGAGAAAAGCAACCCACATGAACTATGATAGCTACAGCGGTGACTTGTCTTAC 1271
QY 345 TyrAspLysAsnIleMetThrLysValIleGlyLysArgTyrAlaTyrLysPheAspPhe 364
Db 1272 TACGATAAAATATATATGATGATTAAGTCCATGCAAGCGCTATGCTTACAAATTTGATTTT 1331
QY 365 HisGlyIleAlaGlnAlaLeuGlnProHisProProGlnuSerSerMetLysTyrPro 384
Db 1332 CATGGGATTGCTCAACGCTTTCACCTCAACCTCCAGAACTCAATTCACCAAGTTCCTCC 1391
QY 385 SerAspLeuProTyrMetSerSerTyrHisAlaHisProGlnuLysMetAsnPheValAla 404
Db 1392 TCAGAACTCCATATATGATGATCATATGATGATGATGATGATGATGATGATGATGATG 1451
QY 405 ProHisProProAlaLeuProValThrSerSerSerPhePheAlaAlaProAsnProTyr 424
Db 1452 CCCCAATCCCGCAGCTTACCTGATGATGATGATGATGATGATGATGATGATGATGATG 1511
QY 425 TrpAsnSerProThrGlyLysIleTyrProAsnThrArgLeuProAlaAlaHisMetPro 444
Db 1512 TGGAAATTCACCACTGGAAGTATTAATCCAAATACCTGGCTGCAAGCTGATGATGCT 1571
QY 445 SerHisLeuGlyThrTyrTrp 451
Db 1572 TCTCATCTTGGAACCTACTAT 1592

RESULT 7
HOMERG2 3166 bp ss-RNA linear PRI 08-NOV-1994
LOCUS Human erg2 gene encoding erg2 protein, complete cds.
DEFINITION M17254
ACCESSION M17254
VERSION M17254.1 GI:182186
KEYWORDS erg 2 protein.
SOURCE Human cell line COLO 320, cDNA to mRNA, clone lambda 12.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 3166)
AUTHORS Rao,V.N., Paps,R.S. and Reddy,E.S.
TITLE erg, a human ets-related gene on chromosome 21: alternative
splicing, polyadenylation, and translation
JOURNAL Science 237 (4815), 635-639 (1987)
MEDLINE 87263429
PUBMED 3299708

FEATURES
source
1..3166
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="21q22.3"
257..1645
/gene="ERG"
257..1645
/gene="ERG"
/note="erg 2 protein"
/codon_start=1

/protein_id="AA52398.1"
/db_xref="GI:182187"
/db_xref="GDB:600-119-884"
translation="MIGVDPDAHIKALSVSBDGLFECAYGTPLHATKEMTASS
SSDYGSTKMSRYRPOQDMLSQPARVITKMECSQVNGNSNDECSVAKGKMG
SPDVTGMVYGSMEKHMPPMTTNERRTVYPADPTLSTDHVOMLEMAVKEGLP
DVNILLFONIGKEIKMTKDDFORLTSYVADLILSLHLYREPLRLHLSVDVKA
LONSRLMHARHTDLPYEPPRSANTGCHPTPOSKAOPSESTYPTKEDORPOLDPY
QILGPTSSRLANPQSGQIQLOWFLLELLSDSSNSCTTWEGTNGEFTKMDPDEVARW
GERKSPMMNVDKLSRALRYDDKNIIMTKRVHGRYAVFDFHGLAQALQHPRESSLY
KYPSDLPYMGSYHAHPDKMNEFVAPHPALPYVTSSSFAPAPNPNYMSPTGATPNTRLP
TSHPMSHLGTYX"
repeat_region 291..298
/note="8-bp repeat sequence"
repeat_region 509..516
/note="8-bp repeat sequence"
polya_signal 3143..3149
/note="putative"
BASE COUNT 915 a 770 c 713 t
ORIGIN 227 bp upstream of PvuII site; chromosome 21.
Alignment Scores:
Pred. No.: 6,696-143 Length: 3166
Score: 2182.00 Matches: 407
Percent Similarity: 91.72% Conservative: 14
Best Local Similarity: 86.67% Mismatches: 24
Query Match: 89.28% Indels: 14
DB: Gaps: 4
US-09-902-772-2 (1-451) x HOMERG2 (1-3166)
QY 2 AlaSerThrIleLysGlnAlaLeuSerValValSerGlnuAspGlnuLeuPheGlnuCys 21
Db 281 GCAGCTCATATCAAGAACCTTATCATGTCAGTGCAGAGACAGTCTGTTGGAGTCT 340
QY 22 AlaTyrGlySerProHisLeuAlaLysThrGlnuMetThrAlaSerSerSerSerGlyTyr 41
Db 341 GCCTACGGAACGCCACACTGCTTAAGACAGATACCCGCTCTCTCCACGACTAT 400
QY 42 GlnuGlnThrSerLysMetSerProArgValProGlnuGlnuAspTrpLeuSerGlnuProPro 61
Db 401 GGACAGACTTCCAAAGATGAGCCACGCGTCCACACAGATGGCTGCTCAACCCCA 460
QY 62 AlaArgValThrIleLysMetGlnuCysAsnProAsnGlnuValAsnGlySerArgAsnSer 81
Db 461 GCGAGGCTCACCATCAAAATGGAATGTAAACCTAGCAGTGAATGGCTCAAGAACTCT 520
QY 82 ProAspAspCysSerValAlaLysGlyGlyLysMetValSerSerSerSerAsnValGly 101
Db 521 CCGATGAATGCAAGTGTGGCCAAAGCGGGAGATGTGGCGCCACCAACACCGTTGG 580
QY 102 MetAsnTyrGlySerTyrMetGlnuGlnuLysHisIleProProProAsnMetThrTrpAsn 121
Db 581 ATGAACTAGGCGACGATCAATGAGGAGGAAAGACATCCACCCCAACATGACACGAAAC 640
QY 122 GlnuArgArgValIleValProAlaAspProThrLeuTrpSerThrAspHisValArgGln 141
Db 641 GAGCGGAGAGTATCTGTCAGAGATCTTACGCTATGAGATACAAACATGTGGCGGAG 700
QY 142 TrpLeuGlnTrpAlaValLysGlnuTyrGlyLysProAsnValAspIleLeuLeuPheGln 161
Db 701 TGGCTGGAATGGCGGTGAAAGATATGGCTTCCAGAGCTCAACATCTTGTATTCAG 760
QY 162 AsnIleAspGlyLysGlnuLeuCysLysMetThrLysAspAspPheGlnuArgLeuTrpPro 181
Db 761 AACATGATGGGAGGAAGATGTGCAAGATGACCAAGAGAGACTCCAGAGGCTCAACCC 820
QY 182 SerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGlnuArgGlyAlaThr 201
Db 821 AGCTACAAAGCCGACATCCTTCTTCACATCTCCACTGAGAGAGCTCCTTCCA 880
QY 202 PheIlePheProAsnThrSerValTyrProGlnuAlaThrGlnuArgIleThrArgPro 221

Db 881 CATTG-----ACTTCA-----GATGATGTGATAAAGCTTACAAAATCTCCCA 925
 QY 222 -----AspleuProTgIUGlnAlaArgArGSerAlaTTP 233
 Db 926 CGGTTAATGATGCTAGAAACACAGATTTCACATATGAGCCCCCAGAGATGACCTCGG 985
 QY 234 ThSerIISerHISProThr---GlnSerIyAlaThrGlnProSerSerSerThVal 252
 Db 986 ACCGGTCACGGCCACCCAGCCCGCCAGTCCGAACCTCCTCAACCATCTCTCCACAGTG 1045
 QY 253 ProIySthrGluAspGlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSer 272
 Db 1046 CCAAAATCTAAGACCGCTCTCTCACTTACATCTTATGATCTTGGACCAACAGT 1105
 QY 273 SerArgLeuAlaAsnProGlySerGlyGlnIleGlnLeuTyrGlnPheLeuGlnLeu 292
 Db 1106 AGCGGCTTGCAATTCAGAGGAGTGGCCAGATCCAGCTTGGCAGTTCCTCTGAGCTC 1165
 QY 293 LeuSerAspSerSerAsnSerAsnCysIleThrTTPoluglyThrAsnGlyGluPheLys 312
 Db 1166 CTGTCGGACAGCTCCACATCTCCAGCTGCATCACCCTGGGAAGGCAACAGGGGAGTTCAG 1225
 QY 313 MetThrAspProAspGluValAlaArgTTPglGluArgIyLysSerLysProAsnMet 332
 Db 1226 ATGACGGATCCGACGAGGTGGCCGGCGCTGGGAGAGCGGAGAGCAACCCACACATG 1285
 QY 333 AsnTyrAspLysLeuSerAlaLeuAlaArgTyrTyrTyrAspLysAsnIleMetThrLys 352
 Db 1286 AACTAGACATAGCTCAGCCGGCCCTCCGTTACTACTATGACAAAGACATCATGACCAG 1345
 QY 353 ValHisGlyLysArgTyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGln 372
 Db 1346 GTCCATGGGAAGCGCTACGGCTACAGTTCAGACTCCACGGAATCGCCAGCCCTCAG 1405
 QY 373 ProHisProGluSerSerMetTyrLysTyrProSerAspLeuProTyrMetSerSer 392
 Db 1406 CCCCACCCCGGAGTATCTCTACAGTACCCCTCAGCTCCCGCTGATGGCTCC 1465
 QY 393 TyrHisAlaHisProGlnLysMetAsnPheValAlaProHisProAlaLeuProVal 412
 Db 1466 TATCAGCCGCCACACAGAAAGATGACTTGTGGCGCCACCTCCAGCCCTCCCGTG 1585
 QY 413 ThrSerSerSerPhePheAlaAlaProAsnProTyrTyrAsnSerProThrGlyIle 432
 Db 1526 ACATCTTCGAGTGTGTGTGGTCCCAACCCATCTGAAATTCACCAACGTGGGGGTATA 1585
 QY 433 TyrProAsnThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr 451
 Db 1586 TACCCCAACACTAGGCTCCCAACCAACCATATGCTTCTCATCTGGGCACTTACTAC 1642
 RESULT 8
 AB031088 1808 bp mRNA linear ROD 07-AUG-2001
 LOCUS AB031088
 DEFINITION Rattus norvegicus VESPI4 mRNA for vascular endothelial cell specific protein 14, complete cds.
 ACCESSION AB031088
 VERSION AB031088.1 GI:15128488
 KEYWORDS vascular endothelial cell specific protein 14.
 SOURCE Rattus norvegicus liver cDNA to mRNA.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathu; Muridae; Murinae; Rattus.
 1 (bases 1 to 1808)
 REFERENCE 1 (bases 1 to 1808)
 AUTHORS Aoki,T., Toyoda,H., Nishimoto,S., Tawara,J., Ukal,Y. and Komuraasaki,T.
 TITLE Identification of VESPI4,a vascular endothelial cell specific protein
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 1808)
 AUTHORS Aoki,T., Toyoda,H., Nishimoto,S., Tawara,J., Ukal,Y. and Komuraasaki,T.
 TITLE Direct Submission

JOURNAL Submitted (12-AUG-1999) Tomomi Aoki, Taisho Pharmaceutical Co., Ltd., Medical Research Laboratories, Molecular Biology Lab., 1-403, Yoshino-cho, Omiya, Saitama 3308530, Japan (E-mail:s17704@cc.taisho.co.jp, Tel:048-663-1111, Fax:048-652-7254)
 FEATURES
 source location/Qualifiers
 1..1808 /organism="Rattus norvegicus"
 /db_xref="taxon:10116"
 /tissue_type="liver"
 1..1808 /gene="VESPI4"
 63..1430 /gene="VESPI4"
 /note="similar to human erg1"
 /product="vascular endothelial cell specific protein 14"
 /protein_id="BAB62744.1"
 /db_xref="GI:15128489"
 /translation="MASTIKELALSVSKDOSLEFCAYGTPHLAKTEMTASSSDYQGT SKSPRYPODDMLSOPTARYTIKMECPNSQVNSRNSPDCSVTKGKMGVSPVGM SVGSYMEKEMPPNMTNERNRIVPADPILMSGDYRDLMAVAKRGYGLDDVILF ONIDKEIKWTDDEDRRLTPSYNADILSHLYRTRPIPLHLSDDYDALQNSPRL MARNRDLPTPEPRRSTWGHSPTRPSKAAQSPSTVPTEDRQPLDYLQIGPTS SRLNPGSGQIQLMQFLLELSDSSNSNCITWEGTNGEFTMDPDEAVARKGERSKP NMNYDKLSRALRYDYDKNIMTKYHGKRYAYKFDHGIQAALQHPPESSLYKYPSDLP YMGSYHTHPQKMFVAPHPALPVPVTSSFPATPNPYNPSPTGIIYPTWRILPASHMPSH Lgityy"
 BASE COUNT 510 a 522 c 426 g 350 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 5,56e-143 Length: 1808
 Score: 2179.00 Matches: 406
 Percent Similarity: 91.30% Conservative: 14
 Best Local Similarity: 88.26% Mismatches: 26
 Query Match: 89.16% Indels: 14
 DB: 10 Gaps: 4
 US-09-902-772-2 (1-451) x AB031088 (1-1808)
 QY 1 MetLaserThrIleLysGluAlaLeuSerValValSerGluAspGlnSerLeuPheGln 20
 Db 63 ATGGCCAGCACTATTAAGAGAGCCCTTGACATTTGAGCAAGACAGATCACTATTGAG 122
 QY 21 CysAlaTyrGlySerProHisLeuAlaLysThrGlnMetThrAlaSerSerSerSerGlu 40
 Db 123 TGTGCTACGGAAAGCCACACCTGGCTAAGACAAAGACCGGATCTCTTCCAGTGAC 182
 QY 41 TyrGlyGlnThrSerLysMetSerProArgValProGlnGlnAspTyrLeuSerGlnPro 60
 Db 183 TATGCCAGACATCCAGATGATAGTCCAGATCCACACAGACATGGCTCTCCAGCCC 242
 QY 61 ProAlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsn 80
 Db 243 CCAGCCAGGGGTACCATCAAGATGAGTGCCTGAGTGAACGAGTTCACAGAAC 302
 QY 81 SerProAspAspCysSerValAlaLysGlyGlyTyrLysMetValSerSerSerAspAsnVal 100
 Db 303 TCACCTGACGATGACAGTGCACCAAGAGTGGGAAGTGGGACGCTGATACGTGTG 362
 QY 101 GlyMetAsnTyrGlySerTyrMetGluGluLysHisIleProProProAsnMetThrThr 120
 Db 363 GGATGAGCTACGGCAGCTACATGAGAGAGACACATGCCGCCCAATATATACACAG 422
 QY 121 AsnGluArgArgValIleValProAlaAspProThrLeuTyrSerThrAspHisValArg 140
 Db 423 AATGAACGACAGAGTATGCTCCCTGCAGATCTCTGTGGACACAGACATGATCGCG 482
 QY 141 GlnTyrLeuGlnTyrPalValValLysGluTyrGlyLeuProAspValAspIleLeuPhe 160
 Db 483 CAGTGGCTGAGTGGGAGTGAAGAAATACGGCTCTTACAGCTGAGACGTCTTATATT 542

QY 161 GlnAsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnAlaGluThr 180
 Db 543 CAGAAATATTGACGAGGAGGAGCTGTGCAGATGACAAAGATACATTCCAGAGGCTACT 602
 QY 181 ProSerTyrAsnAlaAspIleLeuSerHisLeuHisTyrLeuArgLysGlyAla 200
 Db 603 CGSAGTACATGCCAGCAATCTTCTCTCAGATCTCCACTGACAGAGAGACTCCCTT 662
 QY 201 ThrPheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgIleThrArg 220
 Db 663 CCACATCTG-----ACATCC-----GAGCAGCTTGATGAAGCTTACAAACATCTC 707
 QY 221 Pro-----AspLeuProTyrGluGlnIleAlaArgAspSerAla 232
 Db 708 CCACGGTTAATGCATCTTGAACACAGATTTACTTATGAGCTCCACAGAGATCAACC 767
 QY 233 TyrThrSerHisSerHisProThr---GlnSerLysAlaThrGlnProSerSerThr 251
 Db 768 TGGACGGGACAGCAGCCACCCCTCAGTCCAAAGGTCGCCAGCATCTCCCTCCACA 827
 QY 252 ValProLysThrGluAspGlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThr 271
 Db 828 GTGCCCAAACCTGAAACACAGCGCTCCTCAGTTAGATCTTACACAGATCTCGGAGCCACC 887
 QY 272 SerSerArgLeuAlaAsnProGlySerGlyGlnIleGlnLeuTrpGlnPheLeuGlu 291
 Db 888 AGAGTGGCTTGTCTAATCAGAGTACTGGCCAGATCCACTGTGGCAGTTCGTGTAGAA 947
 QY 292 LeuLeuSerAspSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 311
 Db 948 CTCCTCTGTGACAGCTCCCAACTCCATCTGATCTGGAAGGACCAACGGGGAGTTC 1007
 QY 312 LysMetThrAspProAspGluValAlaArgArgTyrGlyGluArgLysSerLysProAsn 331
 Db 1008 AAGATGACAGACCCCGATGAGGTGGCTGGCGGGGAGCGAAGAGCAAGCCCAAC 1067
 QY 332 MetAsnTyrAspLysLeuSerArgAlaLeuArgTyrTyrTyrAspLysAsnIleMetThr 351
 Db 1068 ATGAACTATGACAACTACAGCCGTGCCCTCCCTACTACTACACAAACATCATGACC 1127
 QY 352 LysValHisGlyLysArgTyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeu 371
 Db 1128 AAGGTGCAGGGGAGGAGCTATGCTTCAAGTTGACTTCCAGGGAGTCCCGAGGCGTGG 1187
 QY 372 GlnProHisProProGluSerSerMetTyrLysTyrProSerAspLeuProTyrMetSer 391
 Db 1188 CAGCCCATCCGCCCGAGTGTCTCCCTGTACAAGTACCCCTCCGACCTCGCATATGGGC 1247
 QY 392 SerTyrHisAlaHisProGlnLysMetAsnPheValAlaProHisProProAlaLeuPro 411
 Db 1248 TCTATACACACCCACCCAGAGATGAACTTGTGGCTCCACCCCTCCGGCCCTCCCA 1307
 QY 412 ValThrSerSerSerPhePheAlaAlaProAsnProTyrTyrAsnSerProThrGlyGly 431
 Db 1308 GTCACATCTTCCAGTTCTTGTCTACCCCGAACCCTATGTGAATTCGGCGACTGGGGC 1367
 QY 432 IleTyrProAsnThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyr 451
 Db 1368 AATCAACCCGAACTAGGCTCCAGCCAGCAATATGCCCTTCACCTGGGAGCACTTACTAC 1427
 RESULT 9
 AB073079 2133 bp mRNA linear ROD 17-OCT-2001
 LOCUS DEFINITION Mus musculus Erg mRNA, mouse homolog of human ets-related gene ERG,
 AB073079 transcript variant 2, complete cds.
 ACCESSION AB073079
 VERSION AB073079.1 GI:16197542
 KEYWORDS
 SOURCE Mus musculus CDNA to mRNA, clone: Erg-2.
 ORGANISM Mus musculus
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AUTHORS Ozawa, R., Noguchi, H., Taylor, T. D., Takeda, T., Hattori, M. and Sakaki, Y.
 TITLE Mus musculus Erg mRNA
 JOURNAL Published Only in Database (2001)
 REFERENCE 2 (bases 1 to 2133)
 AUTHORS Ozawa, R., Noguchi, H., Taylor, T. D., Takeda, T., Hattori, M. and Sakaki, Y.
 TITLE Direct Submission
 JOURNAL Submitted (16-OCT-2001) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: hattori@gsc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/, Tel: 81-45-503-9111, Fax: 81-45-503-9170)
 FEATURES
 source location/Qualifiers
 1..2133
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /chromosome="16"
 /clone="Erg-2"
 1..2133
 /gene="Erg"
 149..1540
 /gene="Erg"
 /note="homolog of human ets-related gene (ERG), transcript variant 2"
 /codon_start=1
 /protein_id="BAB69949.1"
 /db_xref="GI:16197543"
 /translation="MIGTVPPDAHIREALSVSDQSLFEACVGTPLHAKTEMTASSSDVGQTKRMSPRPQDDMTSGAPARVITKMCNPSYNGNSRSDSCSYNKGKMGVCSFDYTGKSYGSRMEKRYPPPMNTTNERRVYADPTLMSTDVHQMLEMAWKETGLDNDVLFQNTDKELCNMTKDQDFQRLTFSYNADILSLHYLRTPHLPLSDVDKALQNSPRLMHARNTGGAFLFPMTSVYPAYTORITTPPAQSPSPAPVPTEDQRPQLDYOILPTSSRLANPGSGOIQLOFLEILSDSSNCTIWEETNGEFTMTDPEVARRWGERSKPMNMTDKLSRLRLRYDYDNIMTKVGRKRAYKPDPHGIALQLOHPPESSLYKPSDLPYMSYHNAHPQKMNVSHPHALPALTSSSFASPRPYNSTFGGIYPTRLPASHMPSHLGITTY"
 BASE COUNT 587 a 579 c 531 g 436 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 1,55e-141 Length: 2133
 Score: 2159.50 Matches: 403
 Percent Similarity: 87.00% Conservative: 12
 Best Local Similarity: 84.49% Mismatches: 13
 Query Match: 88.36% Indels: 49
 DB: 10 Gaps: 2
 US-09-902-772-2 (1-451) x AB073079 (1-2133)
 QY 2 AlaSerThrIleLysGlnAlaLeuSerValValSerGluAspGlnSerLeuPheGluCys 21
 Db 173 GCAGCTCATATTAAGAGGCGCTGTGTCAGTTGTGAGCGAGGACAGCAGCTATTGTGAGTGT 232
 QY 22 AlaTyrGlySerProHisIleuAlaLysThrGluMetThrAlaSerSerSerSerGlyTyr 41
 Db 233 GCTTACGAGAACCGCACACCTGTGTAAGACAGATGACCGCATCTCTTCCAGTGTACTAT 292
 QY 42 GLYGlnThrSerLysMetSerProArgValProGlnGlnAspTyrLeuSerGlnProPro 61
 Db 293 GGCCGACATCCAAAGATGATGATGCCAGAGTCCCTCAGCAGAGACTGGGTGTCTCAAGCCCA 352
 QY 62 AlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsnSer 81
 Db 353 GCAGGGGTACACATCAAGATGAGTGCAACCTCAGTCAAGTGAATGTTCCAGGAATCTCA 412
 QY 82 ProAspAspCysSerValAlaLysGlyGlyLysMetValSerSerSerSerAspAsnValGly 101
 Db 413 CCTGATGATGCTCAGTGTGAACAAGGTGGGAAGATGTGGCAGCCCGGATACTGTGGGG 472
 QY 102 MetAsnTyrGlySerTyrMetGluGluLysHisIleProProProAsnMetThrThrAsn 121
 Db 473 ATGAGCTACGGAGCTTACATGAGGAGGAGCATGTGCCCTCCCAATATGACCACAAAAT 532

QY	122	GIUATGATGValIleValIProIlaAspProThrLeuTrpSerThrAspIleValArgIn	141
Db	533	GAGGCGAGATGATCGTCCCTGCGAGATCTTACTCTGTGAGCGACAGACATGTCCGACAG	592
QY	142	TrpLeuGIUTrpAlaValIlysGIUTrpGIULeuProAspValAspIleLeuLeuPheGln	161
Db	593	TGCCTGAGATGGGGGGGAAAGAAATATAGGCGCTCTCGATGTGGAGGTCTTACTATTTCAG	652
QY	162	AsnIleAspLeuLysGIULeuCysLysMetThrLysAspAspPheGlnArgLeuThrPro	181
Db	653	AATATCCATGGGAAGGAGCGTGTCAAGATGACAAAGATGACTTCCAGCGGCTTCACGCG	712
QY	182	SerTrpAsnIleAspIleLeuLeuSerHisIleuNIStyLeuArgGIUArg-----	198
Db	713	AGCTACAAATGCGACATTTCTTCTGCATCTGCATCTCACTACCTCAGAGAGACTCCCTTCCA	772
QY	198	-----	198
Db	773	CATCTGACTTCCGATGACGTGTGAATAGGCTTTACAAACSTCCACGCGTTAATGCATGCC	832
QY	199	-----GIUAlaThrPheIlePheProAsnThrSerValTyTrpProIuAlaThr	214
Db	833	AGAAACACAGGGGTGCGACGCTTTATTATTTCCCAAAATCTTGATATGCCAACCTACAG	892
QY	215	GlnArgIleThrTrpArgProAspLeuProTyGIUArgIleAlaArgSerAlaTrpThr	234
Db	893	CMAAGATTTCMAACTAGGCCA-----	913
QY	235	SerHisSerHisProThrGlnSerLysAlaThrGlnProSerSerSerThrValProLys	254
Db	914	-----GCGTCCACGCCATTCCTCCCTGCAGTGGCCAAA	946
QY	255	ThrIuAspGlnArgProGlnIleuAspProTyGlnIleLeuGlyProThrSerSerArg	274
Db	947	ACTGAACACAGCGCTCTCGATTAGATCTCTTACCAATCTCTGGAGCGACAGTAGCGCG	1006
QY	275	LeuIleAsnProGlySerGIUArgIleGlnLeuTrpGlnPheIleuAsnIleuLeuSer	294
Db	1007	CTTCTAATCCAGGTAGTGGCGCGATGCCAGATCCAGCTGTGGCACTTCTGTCTGCATCTGTCA	1066
QY	295	AspSerSerAsnSerAsnCysIleThrTrpGIUArgIleAsnGIUArgIleuPheLysMetThr	314
Db	1067	GACAGCTCCAACTCCAACTCATCATCTGGGAGGACCAACGCGGAGTTCAAAGATGACA	1126
QY	315	AspProAspGlnValAlaAlaArgTrpGIUArgIleLysSerLysProAsnMetAsnTrp	334
Db	1127	GACCGGACAGAGGTGCTCGCGCTGGGGGAGGAGAAAGCAAGCCCAACTGTGAATAT	1186
QY	335	AspLysLeuSerArgAlaLeuArgTrpTyTrpAspLysAsnIleMetThrLysValHis	354
Db	1187	GACAGGTACGCGCGCGCTCCGCTACTACAGACAAATAATCATGACCAAGGTGCAC	1246
QY	355	GIUArgTrpValArgLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnProHis	374
Db	1247	GGGAGAGCGCTACGCCCTACAAATTTGACTTCCAGGAGATTCGCCAGGCGCTGCAGGCCAC	1306
QY	375	ProProIleuSerMetTyLysTyTrpProSerAspLeuProTyMetLeuSerTrpHis	394
Db	1307	CTCTCTAGTCGTCCCTGTCAAGTACCCCTCCGACTCTCCATACATGGGCTCTCATAC	1366
QY	395	AlaHisProGlnLysMetAsnPheValAlaProHisProProAlaLeuProValHisSer	414
Db	1367	GCCACACCCGAGAAGATGAACTTTGTGTCTCCGCCACCTCCGCGTCTCCCATCACTAC	1426
QY	415	SerSerPhePheAlaAlaProAsnProTyTrpAsnSerProThrLeuGlyIleTyPro	434
Db	1427	TCCAGTTTCTTTGGTCTCCCGCAACCCATCTGGAATTCCACGACTGGGGGCAATCTAC	1486
QY	435	AsnThrArgLeuProAlaAlaHisMetProSerHisLeuGlnThrTyTrp	451
Db	1487	AACCTATGGCTCCCGACACCAATATAGCTCTCTCATCTGGGACCTACTATAC	1537

	RESULT	10				
XLAJ4125	LOCUS	XLAJ4125	1798 bp	mRNA	linear	VRT 21-JAN-2000
	DEFINITION	Xenopus laevis erg gene (erg_A).				
AJ224125	ACCESSION	AJ224125				
AJ224125.1	VERSION	AJ224125.1				
GI:5420045	KEYWORDS	ERG gene; transcription factor.				
	SOURCE	Xenopus laevis.				
	ORGANISM	Xenopus laevis.				
	REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	AUTHORS	Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;				
	TITLE	Xenopodinae; Xenopus.				
	JOURNAL	1 (bases 1 to 1798)				
MEDLINE	PUBMED	Baltzinger,M., Mager-Heckel,A.M. and Remy,P. Xl erg: expression pattern and overexpression during development Dev. Dyn. 216 (4-5), 420-433 (1999)				
REFERENCE	AUTHORS	2 (bases 1 to 1798) Baltzinger,M. Direct Submission Submitted (24-FEB-1998) Baltzinger M., UPR9005, NMDCD, Centre National de la Recherche Scientifique, 15, RUE Rene Descartes, 67084, FRANCE				
FEATURES	source	Location/Qualifiers				
	gene	1..1798 /organism="Xenopus laevis" /db_xref="taxon:8355" /clone="lambda ZapII-erg A" /clone_1bp="lambda ZapI" /dev_stage="stage 30 embryo" 321..1778 /gene="erg" 321..1778 /gene="erg" /function="transcription factor"				
CDS		/codon_start=1 /protein_id="CABA6566.1" /db_xref="GI:5420046" /db_xref="SP:SPREMBL:O9W700" translation="MQMARIPIYEDYSNEALSVSEDOSLPFECTYCTPHLTKEMTNRA SSSDVGYSTMSKPRVPQDWLSOPPSRVTITKMCNPSONGSRSSDDISIGSKRM GGSGDNAGMYNGSYIEKHHSIPMPNTNERIVAPADPTIMSTDHVMQLEMAIKREG LPDVLVLFONIDCKELCKMKTEDEHFILTPSNADIILSHLYREPLPHLDVDV KALONSRLMHARNTGASFEFPNSVYOANOIRIPRODISYSPRSAMTNRAPP SKSQSPVTKPTEPDPPOLDPVOIILGTSRLRAPSCGLOLMOFLLSLSSNNRN CIWEQTNGEPEKATDDPEVARMRERESKPRMAYDKLSRALRYTDKNIRTKVGRRY AYRKDFGLAQLDHPPESTWTKYPSELPLTWSSYVAHPQMFNVADHPALPVTTSSS FFAAPNAVMNSPGISYPNTRLPAASHMSHLGTY"				
BASE COUNT	561 a	429 c	358 g	450 t		
ORIGIN						
Alignment Scores:						
Pred. No.:	3,31e-141	Length:	1798			
Score:	2153.50	Matches:	399			
Percent Similarity:	88.98%	Conservative:	21			
Best local Similarity:	84.53%	Mismatches:	23			
Query Match:	88.11%	Indels:	29			
DB:	5	Gaps:	3			
US-09-902-772-2 (1-451)	x XLAJ4125 (1-1798)					
Oy	7	alaSerValValSerGlunspInserLeupheglucValatyrGlySerPro	26			
Db	366	GAAGCTCTTCAAGTGCGTAGAGATAGTCGTATTTCGAGTGCCACTACGAAAGGCCA	425			
Oy	27	HIsleualalysrthglMetcthrAlaseSerSerSerGluTyrgLyglnThrSerlys	46			
Dd	426	CATCTTACTAAAGCGAATAGCCGCACCTCTCTCATGACTAGGGCAAACCTCGAAA	485			
Oy	47	MeSerProArGvalProGlnGlnAsprTrpleuSerGlnProProAlaArgValthrIle	66			

```

Db      486 ATGAGTCCACGCTGCTTACACAGACTGGCTCTCACAGCTCCATCCAGGGTACCACATC 545
Qy      67 LysMetGluCysAsnProAsnGlnValAsnGlySerArgAsnSerProAspSer 86
Db      546 AAGATGAATGCAACCTTGTCAAGTCATGATCAAGAGAGCTCCCGGATGATCTCAGT 605
Qy      87 ValAlaLysGlyCylLysMetValSerSerSerAspAsnValGlyMetAsnTyrglySer 106
Db      606 ATAGGAAAAGGAAGCAAAATGGGTGGAGGCTCGGATTAATGTGGATTAATGTCACAG 665
Qy      107 TyrMetGluGluLysHisIleProProProAsnMetThrThrasnGlnArgValIle 126
Db      666 TACATAGAAAGAAACACATTTCTCCCAACATGACACATGAGAGAGAGAT 725
Qy      127 ValProAlaAspProThrLeuThrSerThrAspHisValArgGlnTrpLeuGlnTrpAla 146
Db      726 GATACCTGCTGACCTTCTTGTGACACAGATCATGTCAGACAGTGGGTAGAGTGGGA 785
Qy      147 ValLysGluTyrglyLeuProAspValIleLeuLeuPheGlnAsnIleAspGlyLys 166
Db      786 ATAAAAGAAATAGGCGCTTCAGATGCGATGCTCTCCAAAACATCGACGGCAG 845
Qy      167 GluLeuCysLysMetThrLysAspAspPheGlnArgLeuThrProSerTyraAsnAlaAsp 186
Db      846 GATATTCGAAGATGACCAAAAGATTTCCATCAGCTCAGCGCAAGCTACATGCTGAT 905
Qy      187 IleLeuLeuSerHisLeuHisIleTyrglyLeuArgValArg 198
Db      906 ATCCCTCTCGCATCTTCTACCTCAGAGAAACCTCTTCCACATTTGACTTCAGAT 965
Qy      199 -----Gly 199
Db      966 GATGTTGATAGGCGCTTACAAAACCTCCACGGTTAATGCAAGCTAGAAATACAGAGAT 1025
Qy      200 AlaThrPheIlePheProAsnThrSerValTyrglyAlaAlaThrGlnArgIleThrTrh 219
Db      1026 GCCTCTTATTTTCCCAACATCATCCGTTTATCAAGACGCAACCAAGAAATCCCCAGC 1085
Qy      220 ArgProAspLeuProTyrglyGlnAlaArgArgSerAlaTrpThrSerHisSerHisPro 239
Db      1086 AGGCAAGATTTATCATATATGAGCCCTCAGAGAGATGCTGTGGCAATCATCTCGCACCA 1145
Qy      240 ThrGlnSerLysAlaThrGlnProSerSerSerThrValProLysThrGlnAspGlnArg 259
Db      1146 ---CCATCAAAAGCATCTCACCA--TCTACCAAGTTCCAAAACAGAACGCCAGC 1199
Qy      260 ProGlnLeuAspProTyrglyGlnIleLeuGlyProThrSerArgPheAlaAsnProGly 279
Db      1200 CCACAGCTAGATCTTATGATTCCTGGCCACAGACAGCCGACTTGCMAATCCAGCA 1259
Qy      280 SerGlyGlnIleGlnLeuTrpGlnPheLeuLeuGlnLeuLeuSerAspSerAsnSer 299
Db      1260 ACTGGGCAATTCAGCTCGGCAATTCCTTACGGAATCTGGAATCTGCGATAGTCCAACTCC 1319
Qy      300 AsnGlyIleThrTrpGlnLysHisAsnGlyLysPheMetThrAspProAspGluVal 319
Db      1320 AACTGATTACTTGGGAGAGAACCAATGAGAAATTAAGATGACGATCCCGATGAGTG 1379
Qy      320 AlaArgArgTrpGlyLysArgLysSerLysProAsnMetAsnTyraAspLysLeuSerArg 339
Db      1380 GCAAGACGTGGGGGAGAGAGAAACCAACCAACATGACATGACAAAGCTCAGCCGT 1439
Qy      340 AlaLeuArgTyrglyTyraAspLysAsnIleMetThrLysValHisGlyLysArgTyra 359
Db      1440 GCACTTCGTTACTACTACATATAAATATTATGATTAAGTCAATGCAAGCGCTATGCT 1499
Qy      360 TyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnProHisProGluSer 379
Db      1500 TCAAAATTTGATTTTCATGAGATTTGCTCAGGCTTCACCACTCCTCAGATATCAAC 1559
Qy      380 MetTyraLysTyraProSerAspLeuProTyrglyMetSerSerTyraHisAlaHisProGlnLys 399
Db      1560 ATGTACAAGTATCCCTCAGAACTTCATATATGAGTCAATACCATCCACATCCAGAAAG 1619

```

```

Qy      400 MetAsnPheValAlaProHisProProAlaLeuProValThrSerSerPhePheAla 419
Db      1620 ATGAACTTTGTTGCCCCCATCGCCAGCTTACCTGATGACATCGCAAGTTCTTTGCA 1679
Qy      420 AlaProAsnProTyrglyTrpAsnSerProThrGlyGlyIleTyraProAsnThrArgLeuPro 439
Db      1680 GCCCTTAATGCAATACGGAATTCACCACTGGAAGATTTATTCAAATTAATCTGGCTGCA 1739
Qy      440 AlaAlaHisMetProSerHisLeuGlyThrTyrgly 451
Db      1740 GCTAGCATATGCTTCTCTGATCTTGAACCTACTAT 1775

RESULT 11
AB073078      2131 bp      mRNA      linear      ROD 17-OCT-2001
LOCUS      AB073078
DEFINITION      Mus musculus Erg mRNA, mouse homolog of Human ets-related gene ERG,
transcript variant 1, complete cds.
ACCESSION      AB073078
VERSION      AB073078.1 GI:16191716
KEYWORDS
SOURCE
ORGANISM      Mus musculus CDNA to mRNA, clone:Erg-1.
REFERENCE
1      Ozawa,R., Noguchi,H., Taylor,T.D., Takeda,T., Hattori,M. and Sakaki,Y.
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE      Mus musculus Erg mRNA
JOURNAL      Published Only in Database (2001)
REFERENCE      2 (bases 1 to 2131)
AUTHORS      Ozawa,R., Noguchi,H., Taylor,T.D., Takeda,T., Hattori,M. and Sakaki,Y.
FEATURES
source
1..2131
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="16"
/clone="Erg-1"
1..2131
/gene="Erg"
148..1536
/gene="Erg"
/note="homolog of human ets-related gene (ERG), transcript variant 1"
/protein_id="BAB69948.1"
/db_xref="GI:16191717"
/translation="MIGTVPPQPAHIREKALSVSDSLFCAVCTPHILATENTASS
SSDYGQTKMSRPVYQDPMDSQAPRATTKKCNISGVNSRNSPDECSVNGKAWG
SPDYVGMYSIGSEMERHVPVPMNTNERRVIVPADPLTMSDHWQMEAWVEKGLD
DVVDLFFONIGKELECKTKDDFORLTPSYADILSHLHYLRETPDLHLSDYDKA
LONSRLNHAHTDLDPEPRRSATWGSHSLTPQSKAOPSPSAVPTKEDROPDLPY
OLIGPTSSRLANPGSGOIOLOMFLLELSDSNSNCTWEGNGEKKWDDDEVARRV
GERKSRPMNNTDKLSRALRYTYDKIMTKYVSKRRAVYFDHFGTALQALQPHPPSSLY
KPSDLPYMGSTHAHPQKMFVSPHPALPVYSSSFASPMDYMWSPGILYPNTRLP
ASHMPSHLGTY"
BASE COUNT      582 a      590 c      530 g      429 t
ORIGIN
Alignment Scores:
Pred. No.:      6,05e-141      Length:      2131
Score:      2151.00      Matches:      402
Percent Similarity:      91.07%      Conservative:      16
Best Local Similarity:      87.58%      Mismatches:      27
Query Match:      88.01%      Indels:      14
DB:      10      Gaps:      4

```



```

QY 98 AspAsnValGlyMetAsnTyrGlySerTyrMetGluGluLysHisIleProProAsn 117
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 210 GACACCGTTGGATGACTACGACCTACATGAGAGACAGACACACCCCAAC 269
QY 118 MetThrThrAsnGluArgValIleValProAlaAspProThrLeuTyrSerThrAsp 137
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 270 ATGACACGAGACGAGGAGTATCGTCCAGACAGATCCTACGGCATGAGACTACAGAC 329
QY 138 HisValArgIleThrLeuGluTyrPalValValLysGluTyrGlyLeuProAspValAspIle 157
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 330 CAGTCGCGGAGTGGCTGGGCGGTGAAGAAATATGGCTTCACAGCGTCAACATC 389
QY 158 LeuLeuPheGlnAsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGln 177
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 390 TTCCTTTCACAGACATCGATGGAGAGACTGTGCAAGATGACCAAGAGAGACTTCAG 449
QY 178 ArgLeuThrProSerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGlu 197
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 450 AGGCTCACCCCGACGACACAGCCGACATCCTCTCCACATCTCCACATCCTCAGAGAG 509
QY 198 ArgGlyAlaThrPheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgIle 217
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 510 ACTCTCTTCCACATTTG-----ACTTCA-----GATGATGTGATAAGCCTTA 554
QY 218 ThrThrArgPro-----AspLeuProTyrGluGlnAlaArg 229
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 555 CAAACTCTCCACGTTAATGACTGTAGAACACAGATTACCATATGAGCCCCCAGG 614
QY 230 ArgSerAlaThrPheSerHisSerHisProThr---GlnSerLysAlaThrGlnProSer 248
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 615 AGATCAGCGCTGAGCGGTGACAGCCACCCAGCCGACGCGAAGAACTGCTCAACATC 674
QY 249 SerSerThrValProLysThrGluAspGlnArgProGlnLeuAspProTyrGlnIleLeu 268
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 675 CCTTCACAGTGGCCAAACTGAGACACAGCGTCTCAGTACATCCTATCAGATTCTT 734
QY 269 GlyProThrSerSerArgLeuAlaAsnProGlySerGlyGlnIleGlnLeuTyrGlnPhe 288
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 735 GAGACACAGAGTACCGCTTGCAATCCAGGAGTGGCCAGATCCAGCTTGGCAGTTC 794
QY 289 LeuLeuGluLeuLeuSerAspSerSerAsnSerAsnCysIleThrTyrGluGlnIleThrAsn 308
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 795 CTCCTGAGACTCCTGTGCGAGACCTCCAACTCCAGCTGCATCCCGGAGAGCCACCAAC 854
QY 309 GlyGluPheLysMetThrAspProAspGluValAlaArgArgTyrGlyLysSer 328
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 855 GGGGAGTTCAAGATGACGATCCCGACGAGGTGGCCGGCGGTGGGAGAGCGGAAGAGC 914
QY 329 LysProAsnMetAspTyrAspLysLeuSerArgAlaLeuArgTyrTyrAspLysAsn 348
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 915 AAACCCACATGAACTAGATAGCTCAGCCGCGCTCCGTTACTATGACAAAGAAC 974
QY 349 IleMetThrLysValHisGlyLysArgTyrAlaTyrLysPheAspPheHisGlyIleAla 368
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 975 ATCATGACCAAGTCCATGGGAAGCGCTACGCTTACAGTTCACCTCCAGGGATCCGC 1034
QY 369 GlnAlaLeuGlnProHisProProGluSerSerMetTyrLysTyrProSerAspLeuPro 388
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1035 CAGGCCCTCAGGCCACCCCGGAGTATCTGTACAAAGTACCCCTCAGACCTCCG 1094
QY 389 TyrMetSerSerTyrHisAlaHisProGlnLysMetAspPheAlaAlaProHisProPro 408
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1095 TACATGGGCTCCTATACGCCACCCACAGAAAGATAACTTTGTGGCCGCCACCTTCCA 1154
QY 409 AlaLeuProValThrSerSerSerPhePheAlaAlaProAsnProTyrTyrAsnSerPro 428
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1155 GGCCTGCCGCTGACATCTTCAGTTTTCGCGCCCAAAACCATACAGGAATTACCA 1214
QY 429 ThrGlyGlyIleTyrProAsnThrArgLeuProAlaAlaHisMetProSerHisLeuGly 448
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1215 ACTGGGGGTATATACCCAACTAGCTCCACACGACCATATGCTCTCATCTGGGC 1274
QY 449 ThrTyrTyr 451

```

```

Db 1275 ACTTACTAC 1283
RESULT 13
DRE249590
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
1..1436
/organism="Danio rerio"
/db_xref="taxon:7955"
/dev_stage="20-24hr embryo"
45..1400
/gene="fil-1"
45..1400
/gene="fil-1"
/function="transcription factor"
/codon_start=1
/evidence=experimental
/product="fil-1 protein"
/protein_id="CAB56832.1"
/db_xref="GI:600647"
/translation="MDGIIKEALSYSEDSOLFEPYAAAAPLPKTDMTASGTODYQ
THKINPDPQEMINQPYRVNKKREHDINGSRESVQCSGKCKMAGTGEASOMY
TGYMDKCAPPNMTNRRRVIVPADPSLMSPDHYRDLDAIKRYGLOBDTAMFHS
TDGKELCKMSKDDPLRLTSVNTVELSHLNTLRSSSSISITNTPSHADQSRILAKD
DASVDAVRRTGWSNMHSGKSPYVSVSNKPNQPPQDPVQILGPTSSRLANPG
SGQIDLMQFLLELSDSANAGCTWEGNGERKMDPDEVARMRGERSKPMNMDKL
SRALRYVYDKNIMTKVHGKRAYAKDFGIGIQAOLQHPTESTMYKYPSELVYPSYHA
HQOKNVFSPHPSPMVSNSNFEPTTYPWSPPGCIYPRNSVPRHANSYVPSHLSGY
Y"
BASE COUNT 377 a 439 c 349 g 271 t
ORIGIN
Alignment Scores:
Pred. No.: 2..54e-105 Length: 1436
Score: 1637.00 Matches: 303
Percent Similarity: 78.90% Conservative: 56
Best Local Similarity: 66.59% Mismatches: 88
Query Match: 66.98% Indels: 8
DB: 5 Gaps: 5
US-09-902-772-2 (1-451) x DRE249590 (1-1436)
QY 1 MetaLaserThrIleLysGluAlaLeuSerValSerGluAspGlnSerLeuPheGlu 20
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 45 ATGAGCGGAACATATTAAGAGCGGCTGTCAGTGTAGTGAAGACACGATCTGTTCGAG 104
QY 21 CysAlaTyr---GlySerProHisLeuAlaLysThrGluMetThrAlaSerSerSer 39
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

Db 105 CCTCATACGCCGCTGCTCTTTACCCAGACAGCATGCGTCCGGGACTCAG 164
 Qy 40 GtutryglglnthrserlysmetserrproargvalproglnglnasptprrleuSerlgn 59
 Db 165 GATTACGCCAGACGACCAATAATCAACCCCTCCCTTCAGCAAGATGATTATCAG 224
 Qy 60 ProProAlaArgValThrIlelysmetGluCysAsnProAsnGlnValAsnGlySerarg 79
 Db 225 ---CCGGTGGCGGTGAACGTCACCGAGAGTAT-----GACCAATCTCAACGGATCCCA 275
 Qy 80 AsnserProAspAPCysserValAlaIalysgllysmetValIserSerSerAspAsn 99
 Db 276 GAGTCGCCGGGGAGCTGCGAGCGTGGGGAATGCAATTAATGTGGGGGAGCCGAGCCG 335
 Qy 100 ValIglymetAsnTyrglySerTyrmetylglulysHisIleProProAsnmetThr 119
 Db 336 TCTCATGATGAACTATACGGCTGCTGAGAAATGTGCTCCGCCCTCCAAACATGACG 395
 Qy 120 ThrAsnGlnArgArgValIleValProAlaAspProThrleuTppSerThrAspHisVal 139
 Db 396 ACCAATAGAGAGAGTCATCGTCCCGCAGACCCGTCGTGTGTGTCGCGATCACGTCG 455
 Qy 140 ArgGlnThrleuGlnThrPalalaValIleValIleuTyrglyLeuProAspValAspIleleu 159
 Db 456 CGGCAAGTGTAGACTGGCGCATTAAGAGATACGCTCTTCAGAGATCGACAGCCCATG 515
 Qy 160 PheGlnAsnIleAspGlyIleuLeuCysIleuSerThrIlyAspAspPheGlnArgleu 179
 Db 516 TTTTCACGACAGATGTGAAAAAGACCTCTGCMAATGAGCAAAAGACGACTTCTGAGACTC 575
 Qy 180 ThrProserTyraAsnAlaAspIleleuLeuSerHisIleuHisTyrlleuArgGlnArgly 199
 Db 576 ACCAGCCTTATTAACACCGAGTCTTCTCACATCTCAATTAATCTCAGGAAAGTAGC 635
 Qy 200 AlaThrPheIlePheProAsnThrSerValTyrrProGlnAlaThrGlnArgIleThrThr 219
 Db 636 TCATCGATATCATAC--AACAGCCATCTCACGCTGACAGTCCCTCGCTGGCTGCA 692
 Qy 220 ArgProAspLeuProTyrglnGlnAlaArgArgSerAlaThrPrrSerHisSerHisPro 239
 Db 693 AAAAGACATCGTCTTATGATGCTGACGAGGAGCGGCTGCGAACAACATGCACAGC 752
 Qy 240 ThrGlnSerIysAlaThrGlnProSerSerSerThrValProIlyThrGlnAspGlnArg 259
 Db 753 GGAAGAGGCTCTCCAAAGTGGTCTCTCAGAGCGCTGCCAAAAACCCGAGTACAGCTCTGT 812
 Qy 260 ProGlnLeuAspProTyrglnIleleuGlyProThrSerSerArgleuAlaAsnProgly 279
 Db 813 CTTACGCCAATCTCTTATCAGATTTTAGACCCACGACGACTTGGCTTACCCAGGC 872
 Qy 280 SerGlyGlnIleGlnleuTrrPglNpheleuLeuGlnleuLeuSerAspSerSerAsnSer 299
 Db 873 TCAAGTTCAGATCCAGCTGTGGAGTCTCTGTGAGCTCTGTCGACAGCGGCAACGCC 932
 Qy 300 AsnCyrtIleThrPrrGlnGlyThrAsnGlyGlnPheIlysmetThrAspProAspGlnVal 319
 Db 933 GCGTGCATCACCGGGAGAGAACCAAGAGAGTCAAGATGAGAGACCAAGAGAGAGTG 992
 Qy 320 AlaArgTrtpglIygluArglySerIlyProAsnmetAsnTyraPlyIleuSerArg 339
 Db 993 GCAAGGCGGTGGGGCGAAGCAAGAGCAAAACCAACATGAACTAGCAAGAGTACGCCGA 1052
 Qy 340 AlaIleuArgTyrrTyrrTyraPlyAsnIleuIleuThrIlyValHisGlyIlySerArgTyra 359
 Db 1053 GCCCTGGCTACTACTACGACCAAGACATCATGACCAAAATGACGCGCAAGCGCTACGCC 1112
 Qy 360 TyrrIlyAspPheAspPheHisGlyIleAlaGlnAlaIleuGlnProHisProProGlnSerSer 379
 Db 1113 TACAAGTTCAGACTTCCAGGCGATCGCCAGGCGCTGCAACCCCATCCACGAGTCCACC 1172
 Qy 380 MetTyrrIlyTyrrProSerAspLeuProTyrmetserrSerTyrrHisAlaHisProGlnIly 399
 Db 1173 ATGTACAAGTACCAAGAGAGAGTGCCTAGTCTAGCTATACAGCCCAAGAGAG 1232

Qy 400 MetAsnPheValAlaProHisProProAlaLeuProValThrSerSerSerPhePheAla 419
 Db 1233 GTCAACTTCGTCCTCTCACCGCTCCGTCATAGCCCGTACACTCTCTCAACTTCTTCGGT 1292
 Qy 420 AlaProAsnProTyrrTyrrAsnSerProThrIleGlyIleTyrrProAsnThrArgleuPro 439
 Db 1293 CCCACACCCCATACTGTGAGTCCCGCACAGAGAGTATTATCCCAATCCAGAGCTCCCA 1352
 Qy 440 Ala-----AlaHisMetProSerHisIleuGlyThrTyrrTyrr 451
 Db 1353 CGCATGCCAATTACACAGCTGCTTACACACTTGTGATTAAGTACTAC 1397
 RESULT 14
 LOCUS AY029368 1359 bp mRNA linear PRI 09-MAY-2001
 DEFINITION Homo sapiens friend leukemia integration 1 transcription factor
 (FLI1) mRNA, complete cds.
 ACCESSION AY029368
 VERSION AY029368.1 GI:14017402
 KEYWORDS
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens.
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 1359)
 AUTHORS Udhil,B.T.S., Rainey,D.R. and Meredith,D.M.
 TITLE Direct Submission
 JOURNAL Submitted (09-APR-2001) Molecular Medicine Unit, Leeds University,
 Beckett St, Leeds LS9 7TF, United Kingdom
 FEATURES
 source
 1. 1359
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="11"
 1. 1359
 /gene="FLI1"
 1. 1359
 /gene="FLI1"
 /note="ETS transcription factor; involved in the
 pathogenesis of Ewings sarcoma following translocation
 with EMS gene on chromosome 22; may have oncogenic
 potential when fused with EMS gene"
 /codon_start=1
 /product="friend leukemia integration 1 transcription
 factor"
 /protein_id="AAK50443.1"
 /db_xref="GI:14017403"
 /translation="MDGTTKEALVYSDQSLFDSAYGAANHLPKADMTASGSPDYGO
 PHKINLPDPOEWINOPVRYNVREREDHNGSRESPYDCSVSKSLYGGESPNRY
 NSYMDKNGPPPNMTNERRVIVPADPTLMTQEHYOMLEMAIKESIMEIDTSPFO
 NMDGKELCKKMKEDPLRATTLVNTVEVLSLSTYRESLAAYVTTSTDOSSRLSYKE
 DPSYDVRBACWNNNSGINKSPPLGAGATITKNTBQRPPPTDILGFTSSRLANP
 GSGQIQIMQFLLELSDSANASCTTMEGTGERTKMTDPEVARWGERKSPKMNKNDK
 LSRALRTYDKNIMTGVHGRKYAKEDFBIQALQPHPESSMYKPSDISYSPSYH
 AHQQKVFVPPHPSMPVSSSFGAASQYMTSPGTGIYPNPVPRHPNTHVPSHLSS
 Y"
 BASE COUNT 359 a 422 c 333 g 245 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 1.77e-104 Length: 1359
 Score: 1624.50 Matches: 306
 Percent Similarity: 79.43% Conservative: 57
 Best Local Similarity: 66.96% Mismatches: 83
 Query Match: 66.47% Indels: 11
 DB: 9 Gaps: 7
 US-09-902-772-2 (1-451) x AY029368 (1-1359)
 Qy 1 MetAlaSerThrIleIygluAlaLeuSerValIalSerGluAspGlnSerIleuPheGln 20
 Db 1 ATGGACGGAGCTATTAAAGAGGCTGTGTGTGAGCGACGACGACGATCCCTCTTGAC 60

```

QY      21 CysAlaTyrGly---SerProHisLeuAlaLysThrGluMetThrAlaSerSerSer 39
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      61 TCAGCGTACGAGCGAGCGACCCCTTCCTCCCAAGCGCGACATGACTGCTCGGAGTCTCT 120
QY      40 GlnTyrGlnGlnThrSerLysMetSerProAlaValProGlnGlnAspTrpLeuSerGln 59
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      121 GACTACGAGGAGCGCCCAAGATCAACCCCTCCACACGACGAGATGATCAATCAG 180
QY      60 ProProAlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArg 79
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      181 ---CCAGTGAAGGCTCAACGTCGAAGCGGAGTAT-----GACCACATGATGATCAGG 231
QY      80 AsnSerProAspAspCysSerValAlaLysGlyLysMetValSerSerSerAspAsn 99
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      232 GAGTCTCCGCTGAGCTGACGCGCTTACCAATGACGCAAGCTGCTGCGCGAGCGATCC 291
QY      100 ValGlyMetAsnTyrGlySerTyrMetGluGluLysHis---LleProProProAsnMet 118
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      292 AACCCCATGAACTACAAACAGACTATATGACGAGAGAGAAAGGCCCCCTCTCCCAACATG 351
QY      119 ThrThrAsnGlnUarGArgValIleValProAlaAspProThrLeuTrpSerThrAspHis 138
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      352 ACCACCAAGAGAGAGAGTACATGCTCCCGGACGCCACACTGTGGACACAGAGCAT 411
QY      139 ValArgGlnTrpLeuGlnUtrpAlaValLysGlyLysLeuProAspValAlaSpIleLeu 158
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      412 GTGAGCGCAATGCTGAGTGGCGCATAAAGAGTACAGTTGATGAGATCCACATCC 471
QY      159 LeuPheGlnAsnIleAspGlyLysGluLeuCysGlyMetThrLysAspAspPheGlnArg 178
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      472 TTTTTCAGAACTAGATGAGGCAAGAACTGTGTAATATGAAACAGAGGACTTCTCCGC 531
QY      179 LeuThrProSerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGluArg 198
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      532 GCCACACACCTTCAACAAACGAGAGTGTGTGTACACCTCAGTTACCTCAGAGAA--- 588
QY      199 GlyAlaThrPheIlePheProAsnThrSerValTyrProGlnAlaThrGlnArgIleThr 218
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      589 ---AGTTCACTGCTGCCATATATACACCTCCACACGACCAATCCACCATTTGAGT 645
QY      219 ThrArgProAspLeuProTyrGlnGlnAlaArgArgSerAlaTrpThrSerHisSerHis 238
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      646 GTCAAAGAAAGACCTTCTATATGACTCAGTCAGAGAGAGCTGGGGCAATATACATGAT 705
QY      239 ProThrGlnSerLysAlaThrGln---ProSerSerSerThrValProLysThrGluAsp 257
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      706 TCTGGCCTCAACAAAGTCTCTCCCTTGGAGGGGACAAACAGATCAGTAAGATACAGAG 765
QY      258 GlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSerSerArgLeuAlaAsn 277
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      766 CAACGGCCCCAGCAGATCCGATCAGATCTGGGGCCGACGACGAGTGGCCCTAAGCCAAC 825
QY      278 ProGlySerGlyGlnIleGlnLeuTrpGlnPheLeuLeuGlnLeuLeuSerAspSerSer 297
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      826 CCTGGAGAGGGGAGATCCAGCTGTGGCAATTCCTCTGGAGCTGCTCTCCGACAGCGC 885
QY      298 AsnSerAsnCysIleThrTrpGlnGlyThrAsnGlyLysPheLysMetThrAspProAsp 317
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      886 AACGCCAGGTATACATCGGGAGGAGCCAAACGGGAGTCAAAATGACGAGACCCGAT 945
QY      318 GluValAlaLysArgTrpGlyGlnArgLysSerLysProAsnMetAsnTyrAspLysLeu 337
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      946 GAGGTGGCCAGGGGCTGGGGCGAGCGAAAGCAACCCCAACATGAAATACGACAAGCTG 1005
QY      338 SerArgAlaLeuArgTrpTyrTyrAspLysAsnIleMetThrLysValHisGlyLysArg 357
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      1006 AGCGGGGCGCTCGATTATGATGATTAATAAATATTATGACCAATGACGGCAAAAGA 1065
QY      358 TyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnProHisProProGln 377
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      1066 TATGCTTAAATAATTGACTTCCACGGCANTTCCGACAGGCTCTGAGGCACATCCGACCGAG 1125

```

```

QY      378 SerSerMetLysTyrTyrProSerAspLeuProTyrMetSerSerTyrHisAlaHisPro 397
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      1126 TCTGTCATGTACAAAGTACCTTCTGACATCTCTGACATGCTTCTTACACCTGCCACACG 1185
QY      398 GlnLysMetAsnPheValAlaProHisProProAlaLeuProValThrSerSerSerPhe 417
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      1186 CAGAAAGTGACATTTGTCTCCCTCCCATCCATCTCCATGCTCTTCACTTCTCCAGCTTC 1245
QY      418 PheAlaAlaProAsnProTyrTrpAsnSerProThrGlyGlyIleTyrProAsn----- 435
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      1246 TTGTGAGCGCCATCACAATACTGAGCTCCGCCACGAGGGGAGATCTTACCCACCCCAAC 1305
QY      436 ---ThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyr 451
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      1306 GTTCCCCCGCATCTTAAACCCAGTGCCTTTCACACTTACGACGATCTAC 1356

RESULT 15
BC010115
LOCUS      BC010115      2394 bp      mRNA      linear      PRI 12-JUL-2001
DEFINITION Homo sapiens, clone MGC:19589 IMAGE:3635042, mRNA, complete cds.
ACCESSION BC010115
VERSION    BC010115.1 GI:14603315
KEYWORDS   MGC.
SOURCE      Homo sapiens.
ORGANISM   Homo sapiens.
REFERENCE   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS     Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE       1 (bases 1 to 2394)
JOURNAL     Direct Submission
            Submitted (02-JUL-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
REMARK      NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: gcgabs-remail.nih.gov
            Tissue Procurement: Louis W. Staudt, M.D., Ph.D.
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
            DNA Sequencing by: National Institutes of Health Intramural
            Sequencing Center (NISC),
            Galitersburg, Maryland;
            Web site: http://www.nisc.nih.gov/
            Contact: nisc_mgc@ngri.nih.gov
            Shevchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M.,
            Benjamin, B., Blakesley, R.W., Bouffard, G.G., Barkley, C., Brooks, S.,
            Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-T., Karlins, E., Legaspi, R.,
            Lim, M., Maduro, Q.L., Mastello, C., Mastrian, S.D., McCloskey, J.C.,
            McDowell, J., Pearson, R., Snyder, B., Stantrilpop, S., Thomas, P.J.,
            Tjongson, E.E., Touchman, J.W., Tsurgon, C., Vogt, J.L., Walker, M.A.,
            Zhang, L.-H. and Green, E.D.
FEATURES             source
     source
     1..2394
     /organism="Homo sapiens"
     /db_xref="taxon:9606"
     /clone="MGC:19589 IMAGE:3635042"
     /tissue_type="lymph, Burkitt lymphoma"
     /clone_id="NIR_MGC_8"
     /lab_host="DH10B-R"
     /note="vector: pOTB7"
     /codon_start=1
     /product="Unknown (protein for MGC:19589)"
     /protein_id="AAH01015.1"
     /db_xref="GI:14603316"
     /translation="MDGITKEALSVSDQSLFDSAYGAHAHLPRADMTASGSPDYQ

```

PKIKNPLOOEWINOPVNVNRYREYDHNMGSRESPYDSCSKLYGGESNPMY
 NSTYMDKNGPDPNNMTTNERRYVLPADPLMTQEHKRWLEWIKETSLMEIDTSFQ
 NMDKELCKNNKEDRLATLYNTVEVLISLRYLRSLLAYWTISHTDSSSLSYKE
 DPSYDYSVRNAGNNMNSGLNKSPLGAGQTSKNTEORQPPYQILGPTSSRLANP
 GSGOIOLOMOLELLESDSANASCITWEGTGEKMTDPPDPAWRMGKSPMYNDK
 LSRALRYYDKNIMTKYHGRKRYAKRFDFHIOALQPHPRESNMYRPSDSTMPSTH
 AHQOKVNFVPHHSSMPVTSSSFCAASQYTWISPTGIIIPNVNPHNPVHSHLGS
 Y*

BASE COUNT 669 a 609 c 574 g 542 t
 ORIGIN

Alignment Scores:

Pred. No.: 3,47e-104 Length: 2394
 Score: 1624.30 Matches: 306
 Percent Similarity: 79.43% Conservative: 57
 Best Local Similarity: 66.96% Mismatches: 83
 Query Match: 66.47% Indels: 11
 DB: 9 gaps: 7

US-09-902-772-2 (1-451) x BC010115 (1-2394)

QY 1 MetalaserThrllelysglualaleuSerValysergluaspGlnSerLeuPheglu 20
 Db 126 ATGACGCGGACTATTAGAGAGCTCTGCTGAGCCAGCAGCTCCCTTTGAC 185
 QY 21 CysAlatYrgly---SerProhIsLeuAlaLysThrgluMetThrlaLaserSerSer 39
 Db 186 TCAGCGGAGGAGCGGAGCCATCCCAAGCGCAGCATGACTGCTGGGAGAGCT 245
 QY 40 GluTyrTlYlInThrSerLysMetSerProArYValProGlnGlnAspTrpLeuSerGln 59
 Db 246 GACTACGGGAGGAGCCCAAGATCAACCCCTCCACACAGCGAGGATGATCAATCAG 305
 QY 60 ProProAlaArYValThrllelysmetGluCysAnProAsnGlnValAsnGlySerArg 79
 Db 306 ---CCAATGAGGGCTCAACGTCACGGGAGAT-----GACCATGATGATCCAGG 356
 QY 80 AsnSerProAspArYCySerValAlaLysGlyLysMetValSerSerSerAspAsn 99
 Db 357 GAGCTCCGGTGCACGCTGACGTTAGCAAAATGACAGCAAGTGGTGGGAGGAGCTCC 416
 QY 100 ValGlymetAnTrYrglySerTyrMetGluGlnLysHis---lleProProAsnMet 118
 Db 417 AACCCCTGAGACTACAAACACTATATGACGAGAAAGATGCCCTCCCTCCCAACATG 476
 QY 119 ThrThrasnGluArGArValIleValProAlaAspProThrlleuTrpSerThrAspHis 138
 Db 477 ACCACCAAGAGAGAGATGATGCTGCCGACAGACCCACACTGTGTGACACAGAGAGAT 536
 QY 139 ValArgGlnTrpLeuGluTrpAlaValLysGluTyrGlyLeuProAspValAspIleLeu 158
 Db 537 GTGAGGCAATGGCTGAGTGGCCATTAAGAGATACAGCTTGATGATGACACATGCC 596
 QY 159 LeuPheGlnAsnIleAspGlyLysGlnLeuCysLysMetThrlYsAspAspPheGlnArg 178
 Db 597 TTTTCCAGAACATGATGGCAAGAACTGTAAATGAACAAGAGAGACTTCTCCGCG 656
 QY 179 LeuThrProSerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGluArg 198
 Db 657 GCCACCAACCTCTACACAGAGAAAGTGTGTGACACACTGATACCTCAGGAA--- 713
 QY 199 GlyAlaThrPheIlePheProAsnThrSerValTyrProGlnAlaThrglnArgIleThr 218
 Db 714 ---AGTTCACCTGCTGGCCTTAATACAACTCCACACGCAACCAATCCTCAGATGAGT 770
 QY 219 ThrArgProAspLeuProTyrGlnGlnAlaArgSerAlaTrpThrSerHisSerHis 238
 Db 771 GTCAAAAGAAAGACCTTCTTGTGACTCAGTCAGACAGAGAGCTTGGGCAATTAACATGAAT 830
 QY 239 ProThrglnSerLysAlaThrgln---ProSerSerSerThrValProLysThrgluAsp 257
 Db 831 TCTGGCTCAACAAAGTCTCTCCCTTGGAGGGGACAAAGATCAGTAAGATATACAGAG 890

QY 258 GluArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSerSerArgLeuAlaAsn 277
 Db 891 CAAGCGCCCGCCAGCAGATCCGTATCAGATCTCGGCGCCGACAGAGTCCGCTAGCCAAC 950
 QY 278 ProGlySerGlyGlnIleGlnLeuTrpGlnPheLeuLeuGlnLeuSerAspSerSer 297
 Db 951 CCTGAAAGCGGGAGATCCAGCTGTGGCAATCTCTCGGAGCTGCTCTCGACAGCGCC 1010
 QY 298 AsnSerAsnCysIleThrTrpGluGlyThrasnGlyLuphelsmetThraspProasp 317
 Db 1011 AAGCCAGAGCTATCACCCTGGGAGGGGACCAAGCGGAGGTCAAAATGACGAGCCCAT 1070
 QY 318 GluValAlaArGArGTrpGlyGluArgLysSerLysProAsnMetAsnTyrAspLysLeu 337
 Db 1071 GAGGTGGCCAGGCGCTGTGGGCGAGCGGAAAGCAAGCCCAACATGAATACAGCAACCTG 1130
 QY 338 SerArgAlaLeuArGTrpTyrTyrAspLysAsnIleMetThrlYsValHisGlyLysArg 357
 Db 1131 AGCGGCGCTCGCTTATCTATGATMAAACATTATGACCAAAAGTCCAGCGCAAAAGA 1190
 QY 358 TyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnProHisProProGlu 377
 Db 1191 TATGCTTACAAATTTGACTTCCAGGCGATGCCAGGCTCTGCAAGCCACATCCAGCCAG 1250
 QY 378 SerSerMetTyrLysTyrProSerAspLeuProTyrMetSerTyrHisAlaHisPro 397
 Db 1251 TCGTCCATGTCAAGTACCTTGTGACATCTCTACATGCTTCTTACATGCCACAG 1310
 QY 398 GlnYsMetAsnPheValAlaProHisProProAlaLeuProValThrSerSerSerPhe 417
 Db 1311 CAGAAGGGAACCTTGTCCCTCCATCCATCCATCCATCCATCCATCCATCCATCCATCC 1370
 QY 418 PheAlaAlaProAsnProTyrTrpAsnSerProThrglyGlyIleTyrProAsn----- 435
 Db 1371 TTTGGAGCCGATACAAATACTGAGACCTCCCGCCAGCGGGGGAATCTACCCAAACCCCAAC 1430
 QY 436 ---ThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyr 451
 Db 1431 GTCCCGCGCATCTTAACACCCAGCTGCTTACACACTTACAGCAGCTACTAC 1481

Search completed: November 9, 2002, 18:55:08
 Job time : 2945.58 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 9, 2002, 15:37:21 ; Search time 19.9042 Seconds

(with alignments)
939.793 Million cell updates/sec

Title: US-09-902-772-2

Perfect score: 2444
Sequence: 1 MASTIKKALSVSEDSLFEE.....ITPNTKLPAAHPSHLGTYR 451

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	2414.5	98.8	478 1	ERG_CHICK
2	2182	89.3	462 1	ERG_HUMAN
3	1624.5	66.5	452 1	FLI1_HUMAN
4	1613.5	66.0	452 1	FLI1_MOUSE
5	1576	64.5	453 1	FLI1_XENLA
6	1227	50.2	272 1	ERG_MOUSE
7	676	27.7	173 1	ERG_LYTV
8	599	24.5	475 1	ETS6_DROME
9	562.5	23.0	184 1	ETS3_DROME
10	461	18.9	454 1	GABA_MOUSE
11	460	18.8	454 1	GABA_HUMAN
12	454.5	18.6	464 1	ELG_DROME
13	451.5	18.5	438 1	ETLA_XENLA
14	449	18.4	441 1	ETSL_HUMAN
15	448	18.3	440 1	ETSL_MOUSE
16	447	18.3	485 1	ETSL_MOUSE
17	446	18.2	441 1	ETSL_MOUSE
18	446	18.2	441 1	ETSL_MOUSE
19	443.5	18.1	479 1	ETSL_MOUSE
20	437.5	17.9	472 1	ETSL_MOUSE
21	437	17.9	669 1	MYBL_AVILE
22	431.5	17.7	472 1	ETSL_MOUSE
23	429	17.6	469 1	ETSL_MOUSE
24	424.5	17.4	468 1	ETSL_MOUSE
25	379	15.5	718 1	ETSL_MOUSE
26	345	14.1	268 1	ETSL_MOUSE
27	340.5	13.9	551 1	ERG_HUMAN
28	340.5	13.9	551 1	ERG_MOUSE
29	336	13.7	623 1	PNT1_DROME
30	324	13.3	452 1	ETSL_MOUSE
31	321	13.1	110 1	ETSL_MOUSE
32	318.5	13.0	250 1	ETSL_MOUSE
33	317.5	13.0	477 1	ETSL_MOUSE

34	314.5	12.9	477 1	ETV1_MOUSE	P41164 mus musculu
35	308.5	12.6	485 1	ETV2_MOUSE	P97360 mus musculu
36	305	12.5	732 1	POR_DROME	001842 drosophila
37	300.5	12.3	510 1	ERM_HUMAN	P41161 homo sapien
38	298.5	12.2	335 1	ETV2_MOUSE	P41163 mus musculu
39	297.5	12.2	341 1	ETV2_HUMAN	000321 homo sapien
40	292.5	12.0	341 1	TEL2_HUMAN	095603 homo sapien
41	288	11.8	428 1	ELK1_HUMAN	P19419 homo sapien
42	286	11.7	555 1	ETV4_MOUSE	P28322 mus musculu
43	284.5	11.6	430 1	ELK4_MOUSE	P41158 mus musculu
44	283.5	11.6	551 1	ETV4_HUMAN	P43268 homo sapien
45	282.5	11.6	429 1	ELK1_MOUSE	P41969 mus musculu

ALIGNMENTS

RESULT 1
ERG_CHICK ID ERG_CHICK STANDARD: PRT; 478 AA.
AC 090837;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Transcriptional regulator Erg.
GN ERG.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Spleen;
RA MEDLINE=95329425; PubMed=7605748;
RA Dhordain P., Dewitte F., Desbiers X., Stehelin D.,
RA Dutergue-Coquilland M.;
RT "Mesodermal expression of the chicken erg gene associated with
RT precartilaginous condensation and cartilage differentiation";
RL Mech. Dev. 50:17-28(1995).
CC -1- FUNCTION: ACTS AS A TRANSCRIPTIONAL ACTIVATOR.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN MESODERM- AND, TO A LESSER
CC EXTENT, IN ECTODERM-DERIVED TISSUES.
CC -1- SIMILARITY: BELONGS TO THE ETS FAMILY.
CC -1- SIMILARITY: CONTAINS 1 POINTED (PNT) DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
CC or send an email to [license@isb-sid.ch](mailto:license@isb-sib.ch)).

EMBL: X77159; CAA54404.1; -
HSSP: 001543; FLI1.
DR InterPro: IPR000418; Ets.
DR InterPro: IPR002341; HSF_ETS.
DR InterPro: IPR003118; SAM_PNT.
DR Pfam: PF001178; Ets; 1.
DR Pfam: PF02198; SAM_PNT; 1.
DR PRINTS: PR00454; ETSDOMAIN.
DR SMART: SM00251; SAM_PNT; 1.
DR PROSITE: PS00345; ETS_DOMAIN_1; 1.
DR PROSITE: PS00346; ETS_DOMAIN_2; 1.
DR PROSITE: PS0061; ETS_DOMAIN_3; 1.
KW Transcription regulation; Activator; Nuclear protein; DNA-binding.
FT DOMAIN 115 199 POINTED.
FT DNA_BIND 310 390 ETS_DOMAIN.
SQ SEQUENCE 478 AA; 53913 MW; 8DFDB243ED8623A7 CRC64;

Query Match 98.8%; Score 2414.5; DB 1; Length 478;
 Best Local Similarity 94.1%; Pred. No. 6,9e-168;
 Matches 450; Conservative 0; Mismatches 1; Indels 27; Gaps 1;

QY 1 MASTIKALSVSEDSLEFCAYGSPHLAKTEMTASSSEYGGTSMKSPRPVQDMLSDP 60
 DB 1 MASTIKALSVSEDSLEFCAYGSPHLAKTEMTASSSEYGGTSMKSPRPVQDMLSDP 60
 QY 61 PARVITKMECNPNQVNGSRNSPDDCSYAKGKMWSSSDVNGMYGSMYMEKHIPPNNMTT 120
 DB 61 PARVITKMECNPNQVNGSRNSPDDCSYAKGKMWSSSDVNGMYGSMYMEKHIPPNNMTT 120
 QY 121 NERRVIVPADPTLMSDTHVQWLEMAVKEGGLPDVDILLFQNDIGKELCKMTKDDFQRLT 180
 DB 121 NERRVIVPADPTLMSDTHVQWLEMAVKEGGLPDVDILLFQNDIGKELCKMTKDDFQRLT 180
 QY 181 PSYNADILLSHLHYLRER-----GATFIFPNTSYPPA 213
 DB 181 PSYNADILLSHLHYLRER-----GATFIFPNTSYPPA 240
 QY 214 TORITTRPDLPEQARRSAMTSHSPQSKATOPSSSTVPKTEDQRPOLDPYQIIGPTSS 273
 DB 241 TORITTRPDLPEQARRSAMTSHSPQSKATOPSSSTVPKTEDQRPOLDPYQIIGPTSS 300
 QY 274 RLANPGSGOIQIOMQFLELLSDSSNCTTWGCTNGEFTKTPDDEVARRWGKSKPMNN 333
 DB 301 RLANPGSGOIQIOMQFLELLSDSSNCTTWGCTNGEFTKTPDDEVARRWGKSKPMNN 360
 QY 334 YKLSFALRYYYDKNIMTVHGKRYAYKFDGFIQAOLQPHPESSMYRYPSDLPNMSY 393
 DB 361 YKLSFALRYYYDKNIMTVHGKRYAYKFDGFIQAOLQPHPESSMYRYPSDLPNMSY 420
 QY 394 HAHPOKMFVADHPALPVTSSSFPAAPNPYNSPTGGITYPNTRLPAAMPSHLGTY 451
 DB 421 HAHPOKMFVADHPALPVTSSSFPAAPNPYNSPTGGITYPNTRLPAAMPSHLGTY 478

RESULT 2
 ERG_HUMAN STANDARD; PRT; 462 AA.
 ID P11308;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 01-JUL-1989 (Rel. 11, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Transcriptional regulator ERG (Transforming protein ERG).
 GN ERG.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM ERG-2).
 RX MEDLINE=87263429; PubMed=3299708;
 RA Rao V.N., Papas T.S., Shyam E., Reddy P.;
 RT "erg, a human ets-related gene on chromosome 21: alternative splicing,
 RT polyadenylation, and translation.";
 RL Science 237:635-639(1987).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM ERG-1).
 RX MEDLINE=87317608; PubMed=3476934;
 RA Reddy E.S.P., Rao V.N., Papas T.S.;
 RT "The erg gene, a human gene related to the ets oncogene.";
 RL Proc. Natl. Acad. Sci. U.S.A. 84:6131-6135(1987).
 RN [3]
 RP CHROMOSOMAL TRANSLOCATION.
 RX MEDLINE=94356859; PubMed=8076344;
 RA Dunn T., Prassman L., Hages N., Vioja M.V.;
 RT "ERG gene is translocated in an Ewing's sarcoma cell line.";
 RL Cancer Genet. Cytogenet. 76:19-22(1994).
 RN [4]
 RP CHROMOSOMAL TRANSLOCATION.
 RX MEDLINE=94243799; PubMed=8187069;
 RA Ichikawa H., Shimizu K., Hayaashi Y., Ohki M.;

RT "An RNA-binding protein gene, TLS/FUS, is fused to ERG in human
 RT myeloid leukemia with t(16;21) chromosomal translocation.";
 RL Cancer Res. 54:2865-2868(1994).
 CC -1- FUNCTION: ACTS AS A TRANSCRIPTIONAL ACTIVATOR (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: ERG-1 AND ERG-2 (SHOWN HERE);
 CC ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- DISEASE: A FORM OF ACUTE MYELOID LEUKEMIA (AML) IS CHARACTERIZED
 CC BY A CHROMOSOMAL TRANSLOCATION T(16;21)(P11;Q22) WHICH INVOLVES
 CC ERG AND FUS.
 CC -1- SIMILARITY: BELONGS TO THE ETS FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 POINTED (PNT) DOMAIN.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as their content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL, M17254; AAA52398.1; -;
 DR EMBL, M21535; AAA35811.1; -;
 DR PIR, A29515; TVHDEG.
 DR HSP, Q01543; 1EFL.
 DR TRANSFAC; T00265; -;
 DR TRANSFAC; T00266; -;
 DR TRANSFAC; T02129; -;
 DR TRANSFAC; T02130; -;
 DR Genew; HGNC:3446; ERG.
 DR MIM; 165080;
 DR InterPro; IPR000418; Ets.
 DR InterPro; IPR002341; HSE-ETS.
 DR InterPro; IPR003118; SAM_PNT.
 DR Pfam; PF00178; Ets; 1.
 DR Pfam; PF02198; SAM_PNT; 1.
 DR PRINTS; PR00454; ETSDOMAIN.
 DR SMART; SM00413; ETS; 1.
 DR SMART; SM00251; SAM_PNT; 1.
 DR PROSITE; PS00345; ETS_DOMAIN_1; 1.
 DR PROSITE; PS00346; ETS_DOMAIN_2; 1.
 DR PROSITE; PS50061; ETS_DOMAIN_3; 1.
 KW Transcription regulation; Activator; Nuclear protein; DNA-binding;
 KW Alternative splicing; Proto-oncogene; Chromosomal translocation.
 FT DOMAIN 122 206 POINTED.
 FT DNA_BIND 1 99 MISSING (IN ISOFORM ERG-1).
 FT VARSPLIC 294 374
 SQ SEQUENCE 462 AA; 52031 MW; B29F14B0F5C2C697 CRC64;

Query Match 89.3%; Score 2182; DB 1; Length 462;
 Best Local Similarity 88.7%; Pred. No. 4,8e-151;
 Matches 407; Conservative 14; Mismatches 24; Indels 14; Gaps 4;

QY 2 ASITIKALSVSEDSLEFCAYGSPHLAKTEMTASSSEYGGTSMKSPRPVQDMLSDP 61
 DB 9 ASITIKALSVSEDSLEFCAYGSPHLAKTEMTASSSEYGGTSMKSPRPVQDMLSDP 68
 QY 62 ARVITKMECNPNQVNGSRNSPDDCSYAKGKMWSSSDVNGMYGSMYMEKHIPPNNMTT 121
 DB 69 ARVITKMECNPNQVNGSRNSPDDCSYAKGKMWSSSDVNGMYGSMYMEKHIPPNNMTT 128
 QY 122 ERRVIVPADPTLMSDTHVQWLEMAVKEGGLPDVDILLFQNDIGKELCKMTKDDFQRLT 181
 DB 129 ERRVIVPADPTLMSDTHVQWLEMAVKEGGLPDVDILLFQNDIGKELCKMTKDDFQRLT 188
 QY 182 SYNADILLSHLHYLRERGATFFPNTSYPPKATQRTTRP-----DLPEQARRSM 233
 DB 189 SYNADILLSHLHYLRERGATFFPNTSYPPKATQRTTRP-----DLPEQARRSM 243
 QY 234 TSHSPT--OSKATOPSSSTVPKTEDQRPOLDPYQIIGPTSSRLANPGSGOIQIOMQFLEL 292
 DB 244 TSHSPTOSKATOPSSSTVPKTEDQRPOLDPYQIIGPTSSRLANPGSGOIQIOMQFLEL 303

QY 293 LSSNSNCITWEGTNGEFTKTDDEVARWGERKSKPNMNYDKLSRLRYTQKNTMK 352
 DB 304 LSSNSNCITWEGTNGEFTKTDDEVARWGERKSKPNMNYDKLSRLRYTQKNTMK 363
 QY 353 VHGRVYKPFPHGIAALQHPRESSMYKPSDLPYSSVHAHPKKNFAPHPALPV 412
 DB 364 VHGRVYKPFPHGIAALQHPRESSLYKPSDLPYGSYHAPQKNFAPHPALPV 423
 QY 413 TSSSFAPNPYNSPTGIGIPNTRLPAAHPSHLGTY 451
 DB 424 TSSSFAPNPYNSPTGIGIPNTRLPAAHPSHLGTY 462

RESULT 3
 FLII_HUMAN STANDARD: PRT: 452 AA.
 ID FLII_HUMAN 001543: 014319: Q9DE07: 092480;
 AC 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Friend leukemia integration 1 transcription factor (Flt-1 proto-oncogene) (ERGB transcription factor).
 GN FLI1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OC NCBI_TaxID=9606;
 OX 11
 RN SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE=Bone marrow;
 RX MEDLINE=92396239; PubMed=1522903;
 RA Delattre O., Zucman J., Plougastel B., Desmazaie C., Melot T., Peter M., Kovar H., Joubert I., de Jong P., Rouleau G.;
 RT "Gene fusion with an ETS DNA-binding domain caused by chromosome translocation in human tumours.";
 RL Nature 359:162-165(1992).
 RN SEQUENCE FROM N.A. (ISOFORM 1).
 RX MEDLINE=93075640; PubMed=1445800;
 RA Watson D.K., Smyth F.E., Thompson D.M., Cheng J.Q., Testa J.R., Papas T.S., Seth A.;
 RT "The ERG/Flt-1 gene: isolation and characterization of a new member of the family of human ETS transcription factors.";
 RL Cell Growth Differ. 3:705-713(1992).
 RN REVISIONS.
 RA Watson D.;
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 RN SEQUENCE FROM N.A. (ISOFORM 1).
 RX MEDLINE=93007976; PubMed=1394211;
 RA Prasad D.D., Rao V.N., Reddy E.S.;
 RT "Structure and expression of human Flt-1 gene.";
 RL Cancer Res. 52:5833-5837(1992).
 RN SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
 RC TISSUE=Blood;
 RX MEDLINE=93176799; PubMed=8439553;
 RA Thomas R., May W., Denny C., Raskind W., Moore J., Maki R.A., Beck E., Klemasz M.J.;
 RT "Human Flt-1 localizes to chromosome 11q24 and has an aberrant transcript in neuroepithelioma.";
 RL Blochim. Biophys. Acta 1172:155-158(1993).
 RN SEQUENCE FROM N.A. (ISOFORM 1).
 RA Uhl B.T.S., Rainey D.R., Meredith D.M.;
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 RN SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE=Lymph;
 RA Strausberg R.;
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 RN [8]

RP SEQUENCE OF 58-452 FROM N.A.
 RX MEDLINE=98426231; PubMed=9751743;
 RA Zucman-Rossi J., Lecloux P., Victor J.M., Lopez B., Thomas G.;
 RT "Chromosome translocation based on illegitimate recombination in human tumors.";
 RL Proc. Natl. Acad. Sci. U.S.A. 95:11786-11791(1998).
 RN [9]
 RP SEQUENCE OF 278-301 FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=95352541; PubMed=7542907;
 RA Bhadrirath T., Abe S., Nojima T., Yoshida M.C.;
 RT "Molecular analysis of a c(11;22) translocation junction in a case of Ewing's sarcoma.";
 RL Genes Chromosomes Cancer 13:126-132(1995).
 RN [10]
 RP STRUCTURE BY NMR OF 276-373;
 RX MEDLINE=95292091; PubMed=7773776;
 RA Liang H., Mao X., Olejniczak E.T., Nettlesheim D.G., Yu L., Meadows R.P., Thompson C.B., Resik S.W.;
 RT "Solution structure of the ets domain of Flt-1 when bound to DNA.";
 RL Nat. Struct. Biol. 1:871-875(1994).
 CC -1- FUNCTION: SEQUENCE SPECIFIC TRANSCRIPTIONAL ACTIVATOR. RECOGNIZES THE DNA SEQUENCE 5'CA[CA]GGAGT-3'.
 CC -1- SUBUNIT: CAN FORM HOMODIMERS OR HETERODIMERS WITH ETV6/TEL1.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; SEEM TO BE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- DISEASE: A FORM OF EWING'S SARCOMA IS CHARACTERIZED BY A CHROMOSOMAL TRANSLOCATION T(11;22)(Q24;Q12) WHICH INVOLVES FLII AND EWS.
 CC -1- SIMILARITY: BELONGS TO THE ETS FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 POINTED (PNT) DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X67001; CAA47399.1; -;
 DR EMBL; M98833; AAA36812.2; -;
 DR EMBL; S45205; AAB23637.1; -;
 DR EMBL; M93255; AAA58479.1; -;
 DR EMBL; M93255; AAA58480.1; -;
 DR EMBL; AY029368; AAK50443.1; -;
 DR EMBL; BC001670; AAH01670.1; -;
 DR EMBL; BC010115; AAH10115.1; -;
 DR EMBL; Y17293; CAA76731.1; -;
 DR EMBL; D38408; BAA07463.1; ALT_TERM.
 DR PDB; 1FLI; 15-SEP-95.
 DR TRANSLAC: T02066; -;
 DR Genew; HGNC:3749; FLII.
 DR MIM; 193067; -;
 DR InterPro; IPR000418; Ets.
 DR InterPro; IPR002341; HSF_ETs.
 DR InterPro; IPR003118; SAM_PNT.
 DR Pfam; PF00178; Ets; 1.
 DR Pfam; PF02198; SAM_PNT; 1.
 DR PRINTS; PR00454; ETSDOMAIN.
 DR SMART; SM00251; SAM_PNT; 1.
 DR SMART; SM00413; ETS; 1.
 DR PROSITE; PS00345; ETS_DOMAIN_1; 1.
 DR PROSITE; PS00346; ETS_DOMAIN_2; 1.
 DR PROSITE; PS00661; ETS_DOMAIN_3; 1.
 DR Transcription regulation; Activator; DNA-binding; Nuclear protein;
 KW Alternative splicing; Proto-oncogene; Chromosomal translocation;
 KW 3D-structure.
 FT DOMAIN 114 198 POINTED.
 FT DNA_BIND 281 361 ETS-DOMAIN.
 FT VARSPIC 1 76 MDGTRKALSVSDQSIFDSAYGAHLPRADRTASGSPD
 YGQPKINLPQOEWINQPVKVNKREYDHNGS -> ME

```

FT CONFLICT 69 69 GGLAGERA (IN ISOFORM 2).
FT CONFLICT 77 77 E-> V (IN REF. 5).
FT CONFLICT 130 130 MISSING (IN REF. 5).
FT CONFLICT 133 130 P-> A (IN REF. 5).
FT CONFLICT 133 133 W-> V (IN REF. 5).
FT CONFLICT 323 323 E-> Q (IN REF. 4 AND 5).
FT CONFLICT 426 426 MISSING (IN REF. 5).
SQ SEQUENCE 452 AA; 50982 MW; 9C2AAEEBF683F3FA CRC64;

Query Match 66.5%; Score 1624.5; DB 1; Length 452;
Best Local Similarity 67.0%; Pred. No. 1.3e-110;
Matches 306; Conservative 57; Mismatches 83; Indels 11; Gaps 7;

QY 1 MASTIKREALSVSDSLFECAYG-SPLAKTEMTASSSEYQOTSMSTRVPOQMLSQ 59
  1 MDTIKREALSVSDSLFDSATGAANHLPKADMTASGSDYQGPKNINLPQOEWINQ 60
  60 PPARVTIKMECNNOVNSPDCSVAKGKMWSSDVGNNYSYMEKH-IPPNM 118
  61 -PVRVNVKREY--DHNNGSESPVDCSVSKSLVGGGENPANNYSYMEKNGPPPNM 117
  119 TTNERRIVPADPTLWSTDHVROWLEMAVKEGLPDVDILLFONIDGKELCKMTKDDFOR 178
  118 TTNERRIVPADPTLWTOEHVROWLEMAIKEYSLMEIDTSFFONMDGKELCKMKEDFLR 177
  179 LTPSYNADILSHLYLREGCATFIRPNYSVPEAQRITTRDLPYEQARRAMTSHSH 238
  178 ATLYNTEVLLSHLYLRE--SSLLAYNTSHTDQSSRLSVKEDPSYVRGAMNNMN 235
  239 PTQSKAQ--PSSSTVPKTEQRPQLPDYQILGPTSSRLANPGSGQIQLOMFLLELSDSS 297
  236 SGINKSPPLGAGTQISKNTRQRPDPYQILGPTSSRLANPGSGQIQLOMFLLELSDSA 295
  298 NSNCITWEGTNGEFKMTDPEVARRMGERKSKPNMNTDKLSRLRYYYDKNIMTKVHGKR 357
  296 NASCITWEGTNGEFKMTDPEVARRMGERKSKPNMNTDKLSRLRYYYDKNIMTKVHGKR 355
  358 YAKPEFHGAQALOPHPRESSMYKYPBDLPYSSYAHQKKNFAPARHPALPYSSSF 417
  356 YAKPEFHGAQALOPHPRESSMYKYPBDLSYMPHYAHQKKNFAPHPHPSMPTVSSSF 415
  418 FAAPNPYNSPTGIYPN--TRLPAAHMPSHIGTY 451
  416 FGAASQYMTSPITGIYPNVPRHPNTHVSHIGSY 452

RESULT 4
FLIL MOUSE STANDARD: PRT: 452 AA.
ID FLIL MOUSE
AC P26323;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Friend leukemia integration 1 transcription factor (Retroviral
  integration site protein Flil-1).
GN FLIL OR FLI-1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/c; TISSUE-Spleen;
  MEDLINE=91257578; PubMed=2044959;
  Ben-David Y., Giddens E.B., Letwin K., Bernstein A.;
  "Erythroleukemia induction by Friend murine leukemia virus:
  insertional activation of a new member of the ets gene family, flil-1,
  closely linked to c-fets-1.";
  Genes Dev. 5:908-918(1991).
  [2]
  CHARACTERIZATION.
  MEDLINE=93375657; PubMed=8502483;
  Zhang L., Lemarchandel V., Romeo P.-H., Ben-David Y., Greer P.,

```

```

RA Bernstein A.;
RT "The flil-1 proto-oncogene, involved in erythroleukemia and Ewing's
  sarcoma, encodes a transcriptional activator with DNA-binding
  RT specificities distinct from other Ets family members.";
  Oncogene 8:1621-1630(1993).
  CC -1- FUNCTION: SEQUENCE SPECIFIC TRANSCRIPTIONAL ACTIVATOR. RECOGNIZES
  CC THE DNA SEQUENCE 5'C(CA)16GAGCT-3'.
  CC -1- SUBUNIT: CAN FORM HOMODIMERS OR HETERODIMERS WITH ETV6/TEL1 (BY
  CC SIMILARITY).
  CC -1- SUBCELLULAR LOCATION: Nuclear.
  CC -1- TISSUE SPECIFICITY: EXPRESSED IN BOTH HEMATOPOIETIC AND
  CC NONHEMATOPOIETIC TISSUES.
  CC -1- INDUCTION: SPECIFICALLY UP-REGULATED IN LEUKEMIC CLONES WITH
  CC F-MOLV INSERTIONS UP-STREAM OF THE FLI-1 LOCUS.
  CC -1- DISEASE: INVOLVED IN ERYTHROLEUKEMIA INDUCTION BY FRIEND MURINE
  CC LEUKEMIA VIRUS (F-MOLV).
  CC -1- SIMILARITY: BELONGS TO THE ETS FAMILY.
  CC -1- SIMILARITY: CONTAINS 1 POINTED (PNT) DOMAIN.
  CC -----
  CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
  CC the European Bioinformatics Institute. There are no restrictions on its
  CC use by non-profit institutions as long as its content is in no way
  CC modified and this statement is not removed. Usage by and for commercial
  CC entities requires a license agreement (See http://www.isb.ch/announce/
  CC or send an email to license@isb-sib.ch).
  CC -----
  DR EMBL: X59421; CAA42055.1; -.
  DR PIR: S17403; S17403.
  DR HSP: 001543; 1FLI.
  DR TRANSFAC: T01408; -.
  DR MGD: MGI:95554; F111.
  DR InterPro: IPR000418; Ets.
  DR InterPro: IPR002341; HSF_ETs.
  DR InterPro: IPR003118; SAM_PNT.
  DR Pfam: PF001178; Ets; 1.
  DR Pfam: PF02198; SAM_PNT; 1.
  DR PRINTS: PR00454; ETSDOMAIN.
  DR SMART: SM00413; ETS; 1.
  DR SMART: SM00251; SAM_PNT; 1.
  DR PROSITE: PS00345; ETS_DOMAIN_1; 1.
  DR PROSITE: PS00346; ETS_DOMAIN_2; 1.
  DR PROSITE: PS50061; ETS_DOMAIN_3; 1.
  DR Transcription regulation; Activator; DNA-binding; Nuclear protein;
  KW Proto-oncogene.
  FT DOMAIN 114 198 POINTED.
  FT DNA_BIND 281 361 ETS-DOMAIN.
  SQ SEQUENCE 452 AA; 51002 MW; 1F9A06C6893FE2AB CRC64;

Query Match 66.0%; Score 1613.5; DB 1; Length 452;
Best Local Similarity 66.7%; Pred. No. 7.9e-110;
Matches 305; Conservative 56; Mismatches 85; Indels 11; Gaps 7;

QY 1 MASTIKREALSVSDSLFECAYG-SPLAKTEMTASSSEYQOTSMSTRVPOQMLSQ 59
  1 MDTIKREALSVSDSLFDSATGAANHLPKADMTASGSDYQGPKNINLPQOEWINQ 60
  60 PPARVTIKMECNNOVNSPDCSVAKGKMWSSDVGNNYSYMEKH-IPPNM 118
  61 -PVRVNVKREY--DHNNGSESPVDCSVSKSLVGGGENPANNYSYMEKNGPPPNM 117
  119 TTNERRIVPADPTLWSTDHVROWLEMAVKEGLPDVDILLFONIDGKELCKMTKDDFOR 178
  118 TTNERRIVPADPTLWTOEHVROWLEMAIKEYSLMEIDTSFFONMDGKELCKMKEDFLR 177
  179 LTPSYNADILSHLYLREGCATFIRPNYSVPEAQRITTRDLPYEQARRAMTSHSH 238
  178 ATLYNTEVLLSHLYLRE--SSLLAYNTSHTDQSSRLSVKEDPSYVRGAMNNMN 235
  239 PTQSKA--TOPSSSTVPKTEQRPQLPDYQILGPTSSRLANPGSGQIQLOMFLLELSDSS 297
  236 SGINKSPPLGAGTQISKNTRQRPDPYQILGPTSSRLANPGSGQIQLOMFLLELSDSA 295

```

```

OY 298 NSNCITWEGTNGEERKMTDPDEVARWGERKSKPMNNDKLSRALRYYYDKNIMTKVHGKR 357
Db 296 NASCITWGTNGEERKMTDPDEVARWGERKSKPMNNDKLSRALRYYYDKNIMTKVHGKR 355
OY 358 YAKKFDHGIAQALQPHRPRESSMKYPSDLPYMSSTYAHAPQKNFVAPHPALPYTSSSF 417
Db 356 YAKKFDHGIAQALQPHRPRESSMKYPSDLPYMSSTYAHAPQKNFVAPHPALPYTSSSF 415
OY 418 FAAPNPYNSPTGIGIYPN---TRLPAHMPSHLGTYY 451
Db 416 FGAASQYWTSPAGIYVNPSPVPRHPNTHVPSHLASY 452

RESULT 5
FLII_XENLA
ID FLII_XENLA STANDARD: PRT: 453 AA.
AC P4157;
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Retroviral integration site protein FlI-1 homolog.
GN FLI.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94206844; PubMed=8155576;
RA Meyer D., Wolff C.M., Stiegler P., Senan F., Befort N.,
RA Befort J.J., Remy P.;
RT "Xl-flI, the Xenopus homologue of the flI-1 gene, is expressed during
RT embryogenesis in a restricted pattern evocative of neural crest cell
RT distribution."
RL Mech. Dev. 44:109-121(1993).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED DURING EMBRYOGENESIS.
CC -1- SIMILARITY: BELONGS TO THE ETS FAMILY.
CC -1- SIMILARITY: CONTAINS 1 POINTED (PNT) DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X66979; CAA47389.1; -.
DR HSSP; Q01543; IFLI.
DR TRANSFAC; T02067; -.
DR InterPro: IPR000418; ETS.
DR InterPro: IPR002341; HSF_ETS.
DR InterPro: IPR003118; SAM_PNT.
DR Pfam; PF001178; Ets; 1.
DR Pfam; PF02198; SAM_PNT; 1.
DR PRINTS; PR00454; ETSDOMAIN.
DR SMART; SM00413; ETS; 1.
DR SMART; SM00251; SAM_PNT; 1.
DR PROSITE; PS00345; ETS_DOMAIN_1; 1.
DR PROSITE; PS00346; ETS_DOMAIN_2; 1.
DR PROSITE; PS00346; ETS_DOMAIN_3; 1.
DR Transcription regulation; Activator; DNA-binding; Nuclear protein.
KM DOMAIN 113 197 POINTED.
FT DNA BIND 282 362 ETS-DOMAIN.
SQ SEQUENCE 453 AA; 51015 MW; A6E6FCC0EC42D04 CRC64;

Query Match 64.5%; Score 1576; DB 1; Length 453;
Best Local Similarity 66.2%; Pcore. No. 4.2e-107;
Matches 303; Conservative 52; Mismatches 91; Indels 12; Gaps 9;
OY 1 MASTIKALSVSDQSLFECAIG-SPLAKTEMTASSSSEYGTGTSKMSRPVPOODWLQ 59

```

```

Db 1 MDGTTKALSVSDQSLFDSAYGASSHLSKADMTASNPDPGQPHKINP1PPQDDWLQ 60
OY 60 PPARVITMEGNPNQVNSRNSPDCCSYAKGKATVSSSDNVMGMYGMEERH-TPPNM 118
Db 61 -PMRNIRREY--ELMNSRESVPVDCSINKCKLIGSGEGNMMY-TYMDENKNGPPPRM 116
OY 119 TTNERRVIVPADPILWSTDHVQRQWLEMAVKEYGLDPVDILLFONIDKELCKMTKDFOR 178
Db 117 TTNERRVIVPADPILWSTDHVQRQWLEMAVKEYGLVEIDCSLFQONIDKELCKMSKEDFLR 176
OY 179 LTPSNADILLSHLYLEERGATFFPPTSVYPRATORTTRPDLPYQARASAP-TSHS 237
Db 177 STSYTNEVLISHLYNLVLDSSSLGY-WTQAHPTQSSRLFAKEDSYSAVRSNGNMS 235
OY 238 HPTQSKATOPSSSYVPKTED-ORPOLDPYQILGPTSSRLANPGSGOILMOFLLELSDS 296
Db 236 SPVTKSPMGQTQNVNKKSGDOORSDPYQILGPTSSRLANPGSGOILMOFLLELSDS 295
OY 297 SNSCITWEGTNGEERKMTDPDEVARWGERKSKPMNNDKLSRALRYYYDKNIMTKVHGK 356
Db 296 SNSCITWEGTNGEERKMTDPDEVARWGERKSKPMNNDKLSRALRYYYDKNIMTKVHGK 355
OY 357 RYAKKFDHGIAQALQPHRPRESSMKYPSDLPYMSSTYAHAPQKNFVAPHPALPYTSSS 416
Db 356 RYAKKFDHGIAQALQPHRPDTSMYKPSSEFSYMSYSHSHOOKVNFVPSHPSMPTSSG 415
OY 417 FFAAPNPYNSPTGIGIYPN---TRLPAHMPSHLGTYY 451
Db 416 FFGATSPYNSPSANITYPNPVPHPNTHVQSHLOGFY 453

RESULT 6
ERG_MOUSE
ID ERG_MOUSE STANDARD: PRT: 272 AA.
AC P81270;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Transcriptional regulator ERG (Fragment).
GN ERG OR ERG-3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94019387; PubMed=8413305;
RA Rivera R.R., Stulver M.H., Steenbergen R., Murie C.;
RT "Ets proteins: new factors that regulate immunoglobulin heavy-chain
RT gene expression."
RL Mol. Cell. Biol. 13:7163-7169(1993).
CC -1- FUNCTION: ACTS AS A TRANSCRIPTIONAL ACTIVATOR.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE ETS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; S66169; AAB28525.1; -.
DR HSSP; Q01543; IFLI.
DR MGD; MGI:95415; Ery.
DR InterPro: IPR000418; Ets.
DR InterPro: IPR002341; HSF_ETS.
DR Pfam; PF001178; Ets; 1.
DR SMART; SM00413; ETS; 1.
DR PROSITE; PS00345; ETS_DOMAIN_1; 1.
DR PROSITE; PS00346; ETS_DOMAIN_2; 1.

```

DR PROSITE: PS50061; ETS_DOMAIN_3; 1.
 KW Transcription regulation; Activator; Nuclear protein; DNA-binding.
 FT NON_TER 1 1
 FT DNA_BIND 138 218 ETS_DOMAIN.
 FT NON_TER 272 272
 SO SEQUENCE 272 AA; 30907 MW; 895A51C1B14B4A8 CRC64;

Query Match 50.2%; Score 1227; DB 1; Length 272;
 Best Local Similarity 85.3%; Pred. No. 4.4e-82;
 Matches 232; Conservative 2; Mismatches 10; Indels 28; Gaps 2;

OY 174 DDFQRLTPSYNADILSLHYLRER-----GATFRFPN 206
 DB 1 DDFQRLTPSYNADILSLHYLRERPHLTSDVDKALQNSPRLMARNRTGGAATFRPN 60

OY 207 TSYVPATORITTRPDLPEQARRSAMTSHSHT-OSKATQPSSTYVKTEDRPOLDPY 265
 DB 61 TSYVPATORITTRPDLPEQARRSAMTSHSHTLPSKAQSPSAVPKTEDRPOLDPY 120

OY 266 QILGPTSSRLANPGSQIQLMQLLELSDSSNSNCITWEGTNGEERKMDPDDEVARRWG 325
 DB 121 QILGPTSSRLANPGSQIQLMQLLELSDSSNSNCITWEGTNGEERKMDPDDEVARRWG 180

OY 326 RKSKPMNNYKLSRALRYDYDKNIMTKVHGKRRAYKFDHGIQAALQPHRPRESSMYKPS 385
 DB 181 RKSKPMNNYKLSRALRYDYDKNIMTKVHGKRRAYKFDHGIQAALQPHRPRESSLYKPS 240

OY 386 DLPYMSYHAHPQKMFVAPHPALPVTSSSF 417
 DB 241 DLPYMSYHAHPQKMFVAPHPALPVTSSSF 272

RESULT 7
 ERG_LTYVA STANDARD; PRT; 173 AA.
 ID AC 001414;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Transcriptional regulator ERG homolog (Fragment).
 GN ERG.
 OS Lytechinus variegatus (Sea urchin).
 OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
 OC Echinoidae; Euechinoidae; Echinacea; Temnopneustidae;
 OC Lytechinus.
 OX NCBI_Taxid=7654;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93091246; PubMed=1457815;
 RA Qi S., Chen Z.O., Papas T.S., Lautenberger J.A.;
 RT 'The sea urchin erg homolog defines a highly conserved erg-specific
 RT domain.';
 RL DNA Seq. 3:127-129(1992).
 CC -1- FUNCTION: ACTS AS A TRANSCRIPTIONAL ACTIVATOR.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- SIMILARITY: BELONGS TO THE ETS FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: M81067; AAA68905.1; -
 DR HSP: Q01543; 1F1T
 DR InterPro: IPR000418; Ets.
 DR InterPro: IPR002341; HSF_ETs.
 DR Pfam: PF00178; Ets; 1.
 DR SMART: SM00413; ETS; 1.
 DR PROSITE: PS00345; ETS_DOMAIN_1; 1.
 DR PROSITE: PS00346; ETS_DOMAIN_2; 1.

DR PROSITE: PS50061; ETS_DOMAIN_3; 1.
 KW Transcription regulation; Activator; Nuclear protein; DNA-binding.
 FT NON_TER 1 1
 FT DNA_BIND <1 84 ETS_DOMAIN.
 FT NON_TER 173 AA; 19690 MW; F85D496DD58A8F3D CRC64;
 SO SEQUENCE 173 AA; 19690 MW; F85D496DD58A8F3D CRC64;

Query Match 27.7%; Score 676; DB 1; Length 173;
 Best Local Similarity 73.3%; Pred. No. 2.2e-42;
 Matches 129; Conservative 16; Mismatches 23; Indels 8; Gaps 4;

OY 280 SGOIQLMQLLELSDSSNSNCITWEGTNGEERKMDPDDEVARRWGKSKPMNNYKLSR 339
 DB 1 SGOIQLMQLLELSDSSNSNCITWEGTNGEERKMDPDDEVARRWGKSKPMNNYKLSR 60

OY 340 ALRYDYDKNIMTKVHGKRRAYKFDHGIQAALQPHRPRESSMYKPSDLPYMSYHAHPK 399
 DB 61 ALRYDYDKNIMTKVHGKRRAYKFDHGIQAALQMPQVADPMYKQSDLTLYLPET--HPTK 118

OY 400 MNFVAPHPALPVTSSSFPAAPYWNSPYG-GIYPN--TRLPAHMPSHLGTYY 451
 DB 119 INFVGG--TPINPSTNALSLSHSSSYSSPGTANIPSGVHTPHASHMSHIGTYY 172

RESULT 8
 ETS6_DROME STANDARD; PRT; 475 AA.
 ID AC P29776; Q9VP09;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE DNA-binding protein D-ETS-6.
 GN ETS21C OR ETS-6 OR CG2914.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_Taxid=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkeley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brannon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Milos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Frankoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
 RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Meyers A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Galbraith W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Klamel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasmo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Mishina N.V., Moberg C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusslein D.R., Paclet J.M.,
 RA Palazolo M., Pittman G.S., Pan S., Pollard D., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Swislockas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,

RA Williams S.M., Moodage T., Morley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 RN [2]
 RP SEQUENCE OF 235-352 FROM N.A., TISSUE SPECIFICITY, AND DEVELOPMENTAL
 RP STAGE.
 RC STRAIN=Canton-S; TISSUE=Larva;
 RX MEDLINE=92249640; PubMed=1577186;
 RA Chen T., Bunting M., Karim F.D., Thummel C.S.;
 RT "Isolation and characterization of five *Drosophila* genes that encode
 RT an ets-related DNA binding domain.";
 RL Dev. Biol. 151:176-191(1992).
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- TISSUE SPECIFICITY: EMBRYONIC VENTRAL NERVOUS SYSTEM AND 1 PAIR OF
 CC NEURONS IN EACH THORACIC SEGMENT.
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED THROUGHOUT DEVELOPMENT.
 CC -1- SIMILARITY: BELONGS TO THE ETS FAMILY.
 CC -1- CAUTION: Ref.2 sequence differs from that shown due to a
 CC frameshift in position 242.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AE003589; AAF51484.1; -
 DR EMBL; M88475; AAA28452.1; ALT_FRAME.
 DR PIR; S28823; S28823.
 DR HSP; 001543; JPLI.
 DR FLYBase; FBgn0005660; Ets21C.
 DR InterPro; IPR000418; Ets.
 DR InterPro; IPR002341; HSF-ETS.
 DR InterPro; IPR003118; SAM_PNT.
 DR Pfam; PF001178; Ets; 1.
 DR Pfam; PF02198; SAM_PNT; 1.
 DR PRINTS; PR00454; ETSDOMAIN.
 DR SMART; SM00413; ETS; 1.
 DR SMART; SM00251; SAM_PNT; 1.
 DR PROSITE; PS00345; ETS_DOMAIN_1; 1.
 DR PROSITE; PS00346; ETS_DOMAIN_2; 1.
 DR PROSITE; PS50061; ETS_DOMAIN_3; 1.
 KW DNA-binding; Nuclear protein. ETS-DOMAIN.
 FT DNR_BIND 255 335
 FT SEQUENCE 475 AA; 51802 MW; 0D382C41C03B1502 CRC64;
 SQ
 Query Match 24.5%; Score 599; DB 1; Length 475;
 Best Local Similarity 34.6%; Pred. No. 3,2e-36;
 Matches 163; Conservative 46; Mismatches 124; Indels 138; Gaps 20;
 QY 43 QTSMSRRVPO---QDWLSPPARV-----TIKMCNPNQV-NGSRNSPDD 84
 DB 9 QSHQLEPILISQIRSAWCOQRPINAHQDVROKSGISGNETKLAKETEVFTNRRRRR 68
 QY 85 CSYAKGKRWSSSDNVCNMY-----GSYMEEKHI-----PP--- 115
 DB 69 CSSS-----STDSSASSISTSDSGSSSTSSSSINSQDPAALPVPPLIATPTPPAVS 122
 QY 116 -PNNTEERR-----VIVADPTLMSTDHVRQMLEMAVKEYGL-PDVIDLLFQN 162
 DB 123 SPHQAPSPRRSSDSNRSSVSVEYVPPVPHAMTPEDIASWMAKRRKLDPEPIDDFPK 182
 QY 163 IDGEELCKMTKDFQRLTPSTNADILLSHLYLERKATFTFPNTSVYPEAKTORITRPD 222
 DB 183 -DAQELCDISRADPFWCAGSRGGMILAO-HFA-----ISLY-HATGR----- 222
 QY 223 LPYQARARSAMTSHSHPTQSKATQPSSTYVKTEDQRPOLDPQYILGPTSSRLNPGS-G 281

DB 223 -----ETSPMLNDEP--NPYOLINASRLVAQSGCG 253
 QY 282 QIOLMOFLELLSDSSNSNCITWEGTNGCFKMTDPDEVARRRGKRRKNNNYKLSRAL 341
 DB 254 QIOLMOFLELLSDSSNSNAISWEGQSEFRLIDPDEVARRGKRRKNNNYKLSRAL 313
 QY 342 RYTDKNIMTKVHGKRYAKYKDFHGIQALOPH-----PPSSMYKYPSSDLPMSSYNAHP 397
 DB 314 RYTDKNIMTKVHGKRYAKYKDFHGIQALOPH-----PPSSMYKYPSSDLPMSSYNAHP 364
 QY 398 QKMFVAPNHPAL---PYTSSSFRAAPRYNNSPGCIYPTTRLPAAIMP 445
 DB 365 GGAMQGRHPPRLHHNHQSHPHHQLGQPH-----LPHRSSPASNSS 409
 RESULT 9
 ETS3_DROME
 ID ETS3_DROME STANDARD; PRT; 164 AA.
 AC P28774; Q9VRU5; 16-APR-1993 (Rel. 25, Created)
 DT 15-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE DNA-binding protein D-ETS-3.
 GN ETS65A OR ETS3 OR CG7018.
 OS *Drosophila melanogaster* (fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkely;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktarglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhattacharya D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
 RA Burks K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jellali M., Kalush F., Karpen G.H., Ke Z., Kenton J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laske P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Matei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Paclet J.M.,
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D., Schaefer F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Slater E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Sytkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Morley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 RN [2]
 RP SEQUENCE OF 8-69 FROM N.A., FUNCTION, TISSUE SPECIFICITY, AND

```

RP DEVELOPMENTAL STAGE.
RX STRAIN=Canton-S; TISSUE=Larva;
RC MEDLINE=92249640; PubMed=1577186;
RA Chen T., Bunting M., Karim F.D., Thummel C.S.;
RT "Isolation and characterization of five Drosophila genes that encode
RL an ets-related DNA binding domain.";
RN Dev. Biol. 151:176-191(1992).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- TISSUE SPECIFICITY: EMBRYONIC VENTRAL NERVOUS SYSTEM, HIGHER IN
CC THE THORACIC THAN ABDOMINAL SEGMENTS.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED THROUGHOUT DEVELOPMENT.
CC -1- SIMILARITY: BELONGS TO THE ETS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb.ch).
CC -----
DR EMBL; AE003563; AAF50695.1; -
DR EMBL; M88473; AAA28450.2; -
DR PIR; S28820; S28820.
DR HSP; Q01543; 1PRT.
DR Flybase; Fbgn0005658; Ets65A.
DR InterPro; IPR000418; Ets.
DR Pfam; PF00178; Ets; 1.
DR PRINTS; PR00454; ETSDOMAIN.
DR SMART; SM00413; ETS; 1.
DR PROSITE; PS00345; ETS_DOMAIN_1; 1.
DR PROSITE; PS00346; ETS_DOMAIN_2; 1.
DR PROSITE; PS50061; ETS_DOMAIN_3; 1.
DR KW DNA-binding; Nuclear protein.
FT DNA_BIND 11 91 ETS-DOMAIN.
SQ SEQUENCE 184 AA; 20721 MW; 3F8AD7D0C8BCE8B CRC64;

Query Match 23.0%; Score 562.5; DB 1; Length 184;
Best Local Similarity 60.6%; Pred. No. 4,1e-34;
Matches 114; Conservative 25; Mismatches 36; Indels 13; Gaps 7;

QY 268 LGTSSRLANPGSGQTQLOMFLLELSDSSNSCTWEGTNGEFKTDDEVARRRGERK 327
DB 1 MAFSSS-----SSGQIQLOMFLLELSDSSNSCTWEGTNGEFKTDDEVARRRGERK 55
QY 328 SRPNMNYDKLSALRYRYDKNIMTKVHGKRYAKKFDHGIQAOLQPHPPSSMYKYPSDL 387
DB 56 SRPNMNYDKLSALRYRYDKNIMTKVHGKRYAKKFDHGIQAOLQPHPPSSMYKYPSDL 114
QY 388 PFMSSYHAPQKKNVYAPPPALPVTSSTPF--AAPNPNWNSPTGGIY-PN--TRLPAH 442
DB 115 -PMTPIYHSAKLSFWSPPH-HGNTSSASISIFPSAASGWMGNSPATMLXOPHSMSHTVPSH 172
QY 443 MPSHLGTG 450
DB 173 VAPHLSTY 180

RESULT 10
GABA_MOUSE STANDARD; PRT; 454 AA.
AC 000422;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE GA binding protein alpha chain (GABP-alpha subunit).
GN E4Y1A OR GABPA.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RX MEDLINE=91343912; PubMed=1876836;
RA Lamerco K., Thompson C.C., Byers B.P., Walton E.M., McKnight S.L.;
RT "Identification of Ets-and notch-related subunits in GA binding
RL protein.";
RN Science 253:789-792(1991).
CC (2)
CC X-RAY CRYSTALLOGRAPHY (2.15 ANGSTROMS) OF 320-320.
CC MEDLINE=98128030; PubMed=9461436;
CC Batchelor A.H., Payer D.E., de la Brousse F.C., McKnight S.L.,
CC Wolberger C.;
CC "The structure of GABPalpha/beta: an ETS domain-ankyrin repeat
CC heterodimer bound to DNA.";
CC Science 279:1037-1041(1998).
CC -1- FUNCTION: TRANSCRIPTION FACTOR CAPABLE OF INTERACTING WITH PURINE
CC RICH REPEATS (GA REPEATS).
CC -1- SUBUNIT: HETEROETRAMER OF TWO ALPHA AND TWO BETA SUBUNITS.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- TISSUE SPECIFICITY: UBQUITOUS.
CC -1- SIMILARITY: BELONGS TO THE ETS FAMILY.
CC -1- SIMILARITY: CONTAINS 1 POINTED (PNT) DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb.ch).
CC -----
DR EMBL; M74515; AAA53030.1; -
DR PIR; A40858; A40858.
DR PDB; 1AMC; 18-MAR-98.
DR TRANSFAC; T00298; -.
DR TRANSFAC; T01402; -.
DR MGD; MGI:95610; Gabpa.
DR InterPro; IPR000418; Ets.
DR InterPro; IPR002341; HSF_ETTS.
DR InterPro; IPR003118; SAM_PNT.
DR Pfam; PF00178; Ets; 1.
DR Pfam; PF02198; SAM_PNT; 1.
DR PRINTS; PR00454; ETSDOMAIN.
DR SMART; SM00413; ETS; 1.
DR SMART; SM00251; SAM_PNT; 1.
DR PROSITE; PS00345; ETS_DOMAIN_1; 1.
DR PROSITE; PS00346; ETS_DOMAIN_2; 1.
DR PROSITE; PS50061; ETS_DOMAIN_3; 1.
DR KW Transcription regulation; DNA-binding; Nuclear protein; 3D-structure.
FT DOMAIN 170 251 POINTED.
FT DNA_BIND 320 400 ETS-DOMAIN.
SQ SEQUENCE 454 AA; 51363 MW; 76BC47EB2F563AAD CRC64;

Query Match 18.9%; Score 461; DB 1; Length 454;
Best Local Similarity 30.1%; Pred. No. 3,1e-26;
Matches 123; Conservative 64; Mismatches 119; Indels 102; Gaps 14;

QY 15 DQSLFECAYGSPHLAKTEMTASSSEYQGTSKMSPPVPODWLSOPPARVITIMECPNQ 74
DB 78 DRLTFDQ-----VKTDTGVQLSVQVVISYQMEPKL--NILEIYKTAETVEVVIDPD- 127
QY 75 VNGSRNSPDDCSYAKGKAVSSSDNVMNMGYSMEKKHIPPMMNTNER----- 123
DB 128 ---AHHAEEAEHLVEBAQVITLDGT-----KHITTSIDERSSEQVTRMAALEGY 173
QY 124 -----RVIVPADPTLMSTGHVQWLEMAVKEYGLPVDVILLQNTDGRKELCKTKNDPDR 178
DB 174 RKEQERLGIPIYDPIRWSTQOVHLHWVWVKERSMDIDITTL-NISGRRLCSINDQDFQ 232
QY 179 LTPSYNADILLSHLALRRCGAFFIPNTSVYPEATORTITRPDLPYEQARR--SAWTS 236
DB 233 RYV--RGELIMSHLELRK-----YLASQEQOMNEIYTIIDPVQIIPASVVP 278
QY 237 SHPTOSKATQPSSTIVPKTEDQRPQIDPVQIIGPTSSRLANPG-----SGQIQLOMFLLE 291

```

```

Db 279 ATPTTIVINSSAKAA-----KVORSP-RISGEDRS---SPGRRTGNNGOIQLOMQLLE 328
Qy 292 L2SDSSNCTTWEETNGEFKMTDPEVARRRGKSKPMNMYKLSALRYDYDKNMT 351
Db 329 LITDKARDGCSWGDSEFRLNPELVAQKQGRKKPMTNMYKLSALRYDYDKNMT 388
Qy 352 KVHCKRYAYR-----DFHGIAQ 369
Db 389 KVQGRFVYFVCDLTKLIGYSAELNRLVTECQKRLARQMLHGIAQ 436

RESULT 11
GABA_HUMAN STANDARD; PRT; 454 AA.
ID GABA_HUMAN
AC Q06546; Q12939;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE GABA binding protein alpha chain (GABA-alpha subunit) (Transcription factor E4RF1-60) (Nuclear respiratory factor-2 subunit alpha).
GN E4RF1 OR GABPA.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP MEDLINE=93180783; PubMed=8441384;
RA Watanabe H., Sawada J.-I., Yano K.-I., Yamaguchi K., Goto M.,
RT "cDNA cloning of transcription factor E4RF1 subunits with Ets and
RT notch motifs."
RL Mol. Cell. Biol. 13:1385-1391(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=95097980; PubMed=7799916;
RA Guignea S., Vitbastus J.V., Scarpuila R.C.;
RT "Four structurally distinct, non-DNA-binding subunits of human
RT nuclear respiratory factor 2 share a conserved transcriptional
RT activation domain."
RL Mol. Cell. Biol. 15:102-111(1995).
CC -1- FUNCTION: TRANSCRIPTION FACTOR CAPABLE OF INTERACTING WITH PURINE
CC RICH REPEATS (GA REPEATS). NECESSARY FOR THE EXPRESSION OF THE
CC ADENOVIRUS E4 GENE.
CC -1- SUBUNIT: HETEROETRAMER OF TWO ALPHA AND TWO BETA SUBUNITS.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE ETS FAMILY.
CC -1- SIMILARITY: CONTAINS 1 POINTED (PNT) DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D13318; BAA02575.1; -
DR EMBL: U13044; AA65705.1; -
DR HSSP: Q00422; IAWC.
DR TRANSFAC: T01390; -
DR GeneW: HGNC:4071; GABPA.
DR MIM: 600609; -
DR InterPro: IPR0000418; Ets.
DR InterPro: IPR002341; HSF.ETS.
DR InterPro: IPR003118; SAM_PNT.
DR Pfam: PF00178; Ets; 1.
DR Pfam: PF02158; SAM_PNT; 1.
DR PRINTS: PR00454; ETSDOMAIN.
DR SMART: SM00413; ETS; 1.
DR SMART: SM00251; SAM_PNT; 1.
DR PROSITE: PS00345; ETS_DOMAIN_1; 1.

```

```

DR PROSITE: PS00346; ETS_DOMAIN_2; 1.
DR PROSITE: PS50061; ETS_DOMAIN_3; 1.
KW Transcription regulation; DNA-binding; Nuclear protein.
FT DOMAIN 170 251 POINTED.
FT DNA_BIND 320 400 ETS-DOMAIN.
FT CONFLICT 289 290 SS -> RC (IN REF. 2).
FT CONFLICT 440 440 A -> V (IN REF. 2).
SQ SEQUENCE 454 AA; 51295 MW; 1AF2ABBC791910D CRC64;

Query Match 18.8%; Score 460; DB 1; Length 454;
Best Local Similarity 35.5%; Pred. No. 3,7e-26;
Matches 109; Conservative 44; Mismatches 86; Indels 68; Gaps 10;

Qy 111 KHIPPMTTNER-----RYIVPADPTLMSTDHYKQMLENAVKEGLPD 154
Db 150 KHITTSDETSEQYTRMAALEGRKEGERLGIPDPQWSTDDYLHVVMVMEFSMTD 209
Qy 155 VDLIFONIDKEKELCKMTKDFORLTSPYANDLILSHLYREKGAFTFPNTSYPPAT 214
Db 210 IDLTTL-NISGRELCISNQEDFFQRPV--RGEILMSHLELRK-----YVLAS 254
Qy 215 QRTTRDLPYEQARR--SAWTSHPQSKATQPSSTVKTEDQRPQDLPYQLGPTS 272
Db 255 QEQOMNEIWTIDPQVQIIPASVQSATPTTKVIN-SSKAKAKVO-RAPRISGEDRSSP-G 311
Qy 273 SRLANPSGQIOLWQFLLELSDSSNCTTWEETNGEFKMTDPEVARRRGKSKPMM 332
Db 312 NRTGN--NGQQLWQFLLELTDKDARDCSWGDSEFRLNPELVAQKQGRKKRPTM 369
Qy 333 NYDKLSALRYRYDKNIMTKVHGRYAYR----- 362
Db 370 NYEKLSALRYRYDGMICKVQGRFYKFCVCDLTKLIGYSAELNRLVTECQKRLAKM 429
Qy 363 DFHGIAQ 369
Db 430 QLHGIAQ 436

RESULT 12
ELG_DROME STANDARD; PRT; 464 AA.
ID ELG_DROME
AC Q04688; Q9V9A4;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE DNA-binding protein D-ELG.
GN EFS97D OR ELG OR CG6338.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93096481; PubMed=1461651;
RA The S.M., Xie X., Smyth F., Papas T.S., Watson D.K., Schultz R.A.;
RT "Molecular characterization and structural organization of D-Elg, an
RT ets proto-oncogene-related gene of Drosophila."
RL Oncogene 7:2471-2478(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN-Berkeley;
RX MEDLINE=20196006; PubMed=10711132;
RA Adams M.D., Celnikier S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Aghayani A., An H.-J., Andrews-Frankoch C., Baldwin D.,
RA Balow R.M., Basu A., Bayendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhattacharya D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,

```


CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; X52692; CAA36919.1; -
DR PIR; S11225; S11225.
DR HSP; P14921; 2STP.
DR TRANSFAC; T02040; -
DR InterPro; IPR000418; Ets.
DR InterPro; IPR002341; HSF_ETs.
DR InterPro; IPR003118; SAM_PNT.
DR Pfam; PF001178; Ets; 1.
DR Pfam; PF02198; SAM_PNT; 1.
DR PRINTS; PR00454; ETSDOMAIN.
DR SMART; SM00413; ETS; 1.
DR SMART; SM00251; SAM_PNT; 1.
DR PROSITE; PS00345; ETS_DOMAIN_1; 1.
DR PROSITE; PS00346; ETS_DOMAIN_2; 1.
DR PROSITE; PS50061; ETS_DOMAIN_3; 1.
KW DNA-binding; Nuclear protein; Alternative splicing; Phosphorylation.
FT DOMAIN 51 134 POINTED.
FT DNA_BIND 332 412 ETS-DOMAIN.
FT MOD_RES 153 153 PHOSPHORYLATION (BY CAMK) (POTENTIAL).
FT MOD_RES 279 279 PHOSPHORYLATION (BY CAMK) (POTENTIAL).
FT VARSPIC 1 203 MISSING (TN ISOROM C-ETS-1A').
SQ SEQUENCE 438 AA; 50267 MW; B97A8EBDF2DB51FA CRC64;

Query Match 18.5%; Score 451.5; DB 1; Length 438;
Best Local Similarity 31.7%; Pred. No. 1.4e-25;
Matches 119; Conservativity 39; Mismatches 96; Indels 121; Gaps 11;

QY 119 TTNRRVIVADPTLWSTDAVROWLEMAVKEGLPVDILLFOINDKEKCKMTKDFOR 178
DB 55 TKQQRGLGIPDPREWTDMHVRWVAWNEFTLGVDFQK-CMSGAALCALKECFLE 113
QY 179 LTPSYNADILLSHLYRERKQATFIPN-TSVYPEA----- 213
DB 114 LAPDFVDILMEHLEILQNSKQYQSEITPAYPESRYTSDYFISYGLHACVPPSEFS 173
QY 214 -----TQRT-TTRP-----DLEPEQ----- 227
DB 174 EPSRTESYOTLHPISSEELSTKEYNDYPLGLRDLPESLQGDYFTIKQEVVTPDNM 233
QY 228 -----ARSAATSHSHPIQSKATOPSS-----TYPK-----TEQRPQLD 263
DB 234 CLGRISRGKLGQSEFSIESHSCDRLTQSWSSQSYNSIQRPVSYSDFSEDEYPPAMP 293
QY 264 PYQILG-----PTSSRLANGSGOIOLOMOPLELLESDSSNSNCITW 304
DB 294 SHSKSGTFKDYVDRALNDRKPYTPAALAGTYGSGPIQLOMOPLELLELDKSCQSTISW 353
QY 305 EGTNGEFTKMDPDEVARRMGRKSKPMNMTDKLSRALRYYYKKNIMTKVGRKRAYKE-- 362
DB 354 TGDQMEFKLSDPDDEVARRMGRKPKMNYEKLISRLRYYYDKNIIKHTAGKRRVYRVC 413
QY 363 DFHGINAQLPHPE 377
DB 414 DL-----OSLLGYPE 424

RESULT 14
ETSL_HUMAN
ID ETSL_HUMAN STANDARD; PRT; 441 AA.
AC P14921;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE C-ets-1 protein (p54).
GN ETSL.

OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89083219; PubMed=3060801;
RA Reddy E.S.P., Rao V.N.;
RT "Structure, expression and alternative splicing of the human c-ets-1
RT proto-oncogene";
RL Oncogene Res. 3:239-246(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=89042086; PubMed=2847145;
RA Watson D.K., McWilliams M.J., Lapis P., Lautenberger J.A.,
RA Schweinfest C.W., Papas T.S.;
RT "Mammalian ets-1 and ets-2 genes encode highly conserved proteins";
RL Proc. Natl. Acad. Sci. U.S.A. 85:7862-7866(1988).
RN [3]
RP STRUCTURE BY NMR OF 320-415.
RX MEDLINE=96097120; PubMed=8521493;
RA Werner M.H., Clore G.M., Fisher C.L., Fisher R.J., Trinh L.,
RA Shiloach J., Gronenborn A.M.;
RT "The solution structure of the human ETSL-DNA complex reveals a novel
RT mode of binding and true side chain intercalation";
RL Cell 83:761-771(1995).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOROMS: C-ETS-1A (SHOWN HERE) AND C-ETS-
CC 1B; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- PTM: ISOFORMS OF ETS-1 ARE ALSO GENERATED BY DIFFERENTIAL
CC PHOSPHORYLATION.
CC -1- DISEASE: ETS IS RESPONSIBLE FOR ERYTHROBLAST AND FIBROBLAST
CC TRANSFORMATION. THE JUXTAPosition OF THE INTERFERON AND C-ETS-1
CC PROTO-ONCOGENE MAY BE INVOLVED IN THE PATHOGENESIS OF HUMAN
CC MONOCYTIC LEUKEMIA.
CC -1- SIMILARITY: BELONGS TO THE ETS FAMILY.
CC -1- SIMILARITY: CONTAINS 1 POINTED (PNT) DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; X14798; CAA32904.1; -
DR EMBL; X14798; CAA32903.1; -
DR EMBL; J04101; AAA52410.1; -
DR PIR; A32066; TVHUET.
DR PIR; S10086; S10086.
DR PDB; 2STP; 12-MAR-97.
DR PDB; 2STW; 12-MAR-97.
DR TRANSFAC; T00112; -
DR TRANSFAC; T01400; -
DR Transf; HGNC:3488; ETSL.
DR MIM; 164720; -
DR InterPro; IPR000418; Ets.
DR InterPro; IPR002341; HSF_ETs.
DR InterPro; IPR003118; SAM_PNT.
DR Pfam; PF001178; Ets; 1.
DR Pfam; PF02198; SAM_PNT; 1.
DR PRINTS; PR00454; ETSDOMAIN.
DR SMART; SM00251; SAM_PNT; 1.
DR SMART; SM00413; ETS; 1.
DR PROSITE; PS00345; ETS_DOMAIN_1; 1.
DR PROSITE; PS00346; ETS_DOMAIN_2; 1.
DR PROSITE; PS50061; ETS_DOMAIN_3; 1.
KW Proto-oncogene; DNA-binding; Nuclear protein; Alternative splicing;
KW Phosphorylation; 3D-structure.
FT DOMAIN 53 136 POINTED.
FT DNA_BIND 335 415 ETS-DOMAIN.
FT MOD_RES 38 38 PHOSPHORYLATION (BY MAPK) (BY

[illegible]

RN		baculovirus expression system.";
RL	Oncogene Res.	5:277-285(1990).
RA	[4]	
RB	STRUCTURE BY NMR OF 29-138.	
RC	MEDLINE=98445336; PubMed=9770451;	
RD	Slusky J.C.M., Gentile L.N., Donaldson L.W., Mackereith C.D.,	
RE	Seidel J.J., Graves B.J., McIntosh L.P.;	
RF	"Structure of the ets-1 pointed domain and mitogen-activated protein	
RG	kinase phosphorylation site."	
RH	Proc. Natl. Acad. Sci. U.S.A.	95:12129-12134(1998).
RI	[5]	
RJ	STRUCTURE BY NMR OF 332-415.	
RK	MEDLINE=96176767; PubMed=8598195;	
RL	Donaldson L.W., Petersen J.M., Graves B.J., McIntosh L.P.;	
RM	"Solution structure of the Ets domain from murine Ets-1: a winged	
RN	helix-turn-helix DNA binding motif."	
RO	EMBO J.	15:125-134(1996).
RP	-1 SUBCELLULAR LOCATION: Nuclear.	
RQ	-1 SIMILARITY: BELONGS TO THE ETS FAMILY.	
RS	-1 SIMILARITY: CONTAINS 1 POINTED (PNT) DOMAIN.	
RT	-----	
RU	This SWISS-PROT entry is copyright. It is produced through a collaboration	
RV	between the Swiss Institute of Bioinformatics and the EMBL Outstation -	
RW	the European Bioinformatics Institute. There are no restrictions on its	
RX	use by non-profit institutions as long as its content is in no way	
RY	modified and this statement is not removed. Usage by and for commercial	
RZ	entities requires a license agreement (See http://www.isb.ch/announce/	
SA	or send an email to license@isb-sib.ch).	
SB	-----	
SC	EMBL; M58482; AAA63299.1; -	
SD	EMBL; X53953; CAA37904.1; -	
SE	EMBL; X55787; CAA39310.1; -	
SF	PIR; A30487; A30487.	
SG	PDB; 1ETC; 29-JAN-96.	
SH	PDB; 1ETD; 29-JAN-96.	
SI	PDB; 1BOV; 18-NOV-98.	
SJ	TRANSPAC; T00111; -	
SK	MCD; MGI; 95455; Ets1.	
SL	InterPro; IPR000418; Ets.	
SM	InterPro; IPR002341; HSF_ET5.	
SN	InterPro; IPR003118; SAM_PNT.	
SO	Pfam; PF00178; Ets; 1.	
SP	Pfam; PF02198; SAM_PNT; 1.	
SQ	PRINTS; PR00454; ETSDOMAIN.	
SR	SMART; SM00413; ETS; 1.	
SS	SMART; SM00251; SAM_PNT; 1.	
ST	DR PROSITE; PS00345; ETS_DOMAIN_1; 1.	
SV	DR PROSITE; PS00346; ETS_DOMAIN_2; 1.	
SW	DR PROSITE; PSS0061; ETS_DOMAIN_3; 1.	
SX	KW Photo-oncogene; DNA-binding; Nuclear protein; Alternative splicing;	
SY	Phosphorylation; 3d-structure.	
SZ	DOMAIN 53 136	
TA	POINTED.	
TB	ETS-DOMAIN.	
TC	PHOSPHORYLATION (BY MAPK).	
TD	D->E (IN REF. 1 AND 3).	
TE	L->S (IN REF. 3).	
TF	AT->SY (IN REF. 3).	
TG	G->P (IN REF. 1).	
TH	L->R (IN REF. 3).	
TI	Q->R (IN REF. 3).	
TJ	D->E (IN REF. 3).	
TK	L->V (IN REF. 3).	
TL	D->V (IN REF. 3).	
TM	Q->H (IN REF. 3).	
TN	E->D (IN REF. 3).	
TO	H->R (IN REF. 3).	
TP	A->R (IN REF. 3).	
TQ	D->N (IN REF. 3).	
TR	G->C (IN REF. 3).	
TS	K->S (IN REF. 3).	
TT	G->A (IN REF. 3).	
TU	KR->NA (IN REF. 3).	
TV	R->A (IN REF. 3).	
TW	SEQUENCE 440 AA; 50201 MW; 151164D83CA1BI43 CRC64;	

Query Match 18.3%; Score 448; DB 1; Length 440;
Best Local Similarity 30.6%; Pred. No. 2,6e-25;
Matches 115; Conservative 45; Mismatches 94; Indels 122; Gaps 9;

```
QY 119 TTNRRRIYVADPIIMSTDIHVRQMLEWAVKEGLPDVILLFONIDGKEICKMTKDDFOR 178
      |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 57 TKEQQRIGIPDPDPRQWTEHVDMWMAVNEFSLKGVDFQKF-CMSGALCALGKECFLE 115
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 179 LTPSYNADILLSHLHYLRE-----RGATFIRPNT----- 207
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 116 LAPDFVGDILMEHLEILQKEVAKRYQVNGANPTYPESCITSYRISYIEHAQCVPSEF 175
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 208 -----SYYPEA----- 222
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 176 SEPSFTESYQTLHPISSEELLSLKENDYPSVILQDPLQDTLQTDYFAIKQEVLPDN 235
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 223 LPTYQARR-----SMTSHSHPTQSKATQPSSTVPEK-----ED----- 257
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 236 MCLGRASRGKLGQDSFESVESYDSCDRLTQSWSSQSFNSLQRPVPSYDSFDYEDYPAL 295
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 258 -----QRPQLDPYQILGPTSSRIANPGSGQIQLOMQLLELLSDSSNSNCIT 303
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 296 PNHKPKGTFFKDYVDRADLNKDFVTPAALAGTGGPIQLMQLLELLTDKSCQSFIS 355
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 304 WEGTNGEFKMTDPDEVARRMGERKSKPMNYDKLSRALRYYYDKNIMTKYHGKRYAYKF- 362
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 356 WTGDMFEFKLSDPDEVARRMGKRNKPKMNYEKLSRGLRYYYDKNIIHKTAGKRYVRFV 415
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 363 -DFHGIQAQLQPHPE 377
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 416 CDL---QSLGLYPE 427
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
```

Search completed: November 9, 2002, 16:51:17
Job time : 21.9042 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 9, 2002, 16:54:47 ; Search time 225.257 Seconds

(without alignments)
4508.850 Million cell updates/sec

Title: US-09-902-772-2

Perfect score: 2444
Sequence: 1 MASTREALSVSEDSLFE.....ITPNTRLPAHMPSHLGYTY 451

Scoring table:

BIOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2185239 segs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlp
-O=/cgn2.1/USPTO.spool/US09902772/rnat.07112002.134033.9021/app.query.fasta_1.1294
-DB=N.Geneseq.101002 -OPMT=fastap -SUFFIX=mg -MINMATCH=0.1 -LOOPTC=0
-LOOPTXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdl
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pcit -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09902772.eccn.1_1.103.etrnat.07112002.134033.9021 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-MARN_TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: N.Geneseq.101002.*

1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	2444	100.0	1447	20	AAAX26551
2	2414.5	98.8	1528	20	AAAX26552
3	1624.5	66.5	2938	14	AAO50644
4	1624.5	66.5	2957	24	ABR84139
5	1599	65.4	2954	22	AAH02915
6	1506.5	61.6	1890	14	AAO50652
7	948	38.8	567	22	ABA48124
8	948	38.8	567	22	ABA66003
9	948	38.8	567	22	ABA33090
10	948	38.8	567	22	AAK44426
11	948	38.8	567	22	AAK40160
12	948	38.8	567	22	AAI20992
13	948	38.8	567	22	AAI46176
14	948	38.8	567	22	AAI06643
15	948	38.8	567	24	ABS14215
16	863	35.3	473	22	ABA50193
17	863	35.3	473	22	ABA68128
18	863	35.3	473	22	ABA35152
19	863	35.3	473	22	AAK16512
20	863	35.3	473	22	AAK42265
21	863	35.3	473	22	AAI48339
22	863	35.3	473	22	AAI08696
23	712	29.1	549	24	ABR44114
24	654	26.8	533	21	AAO01342
25	650	26.6	420	22	ABA45056
26	650	26.6	420	22	ABA55529
27	650	26.6	420	22	ABA52539
28	650	26.6	420	22	AAK03760
29	650	26.6	420	22	AAK29224
30	650	26.6	420	22	AAI35180
31	650	26.6	420	22	AAI03692
32	650	26.6	420	22	ABA42995
33	650	26.6	420	22	ABA53411
34	650	26.6	420	22	ABA23187
35	650	26.6	420	22	AAK01681
36	650	26.6	420	22	AAK27129
37	650	26.6	420	22	AAI11718
38	650	26.6	420	22	AAI33025
39	650	26.6	420	22	AAI01646
40	650	26.6	420	22	ABO501685
41	599.5	24.5	2025	23	ABL05371
42	562.5	23.0	899	23	ABL12097
43	511	20.9	1752	22	AAI3672
44	511	20.9	1752	24	AAI38753
45	498	20.4	318	21	AAI21875

ALIGNMENTS

RESULT 1
AAAX26551
ID AAAX26551 standard; DNA: 1447 BP.

AC AAAX26551:

DT 14-JUN-1999 (first entry)

DE DNA encoding chicken C-11 protein.

KW Chicken; C-11 protein; cell calcification inhibiting activity;

KW cell calcification inhibiting agent; c-erg protein; arthritis deformans;

OS Gallus sp.

PH Key Location/Qualifiers
CDS 63..1418

KW ossification: spinal column ligament: ss.
 XX Gallus sp.
 OS
 XX Key Location/Qualifiers
 FH CDS 63.1499
 FT /*tag= a
 XX
 XX JP11075871-A.
 XX
 XX 23-MAR-1999.
 PD
 XX
 XX 29-MAY-1998; 98BP-0166076.
 PF
 XX 20-JUN-1997; 97US-0050297.
 PR 18-JUN-1997; 97US-0878177.
 XX
 PA (CHUS) CHUGAI PHARM CO LTD.
 PA (UTPE-) UNIV PENNSYLVANIA.
 XX
 DR WPI; 1999-257708/22.
 DR P-PSDB; AAY01521.
 XX
 XX
 PT An active protein for inhibiting cell calcification - useful for
 PT measuring the calcification of a cell, for diagnosing arthritis
 PT deformans or ossification of spinal column ligament
 XX
 XX Disclosure; Page 8-9; 15pp; Japanese.
 XX
 CC The present sequence encodes a chicken c-ery protein. The specification
 CC also describes a chicken C-11 protein (AA01520) which has cell
 CC calcification inhibiting activity and a cell calcification inhibiting
 CC agent containing c-ery protein. The proteins are used for measuring the
 CC calcification of a cell, for diagnosing arthritis deformans or
 CC ossification of spinal column ligament.
 XX
 SQ Sequence 1528 BP; 464 A; 395 C; 330 G; 339 T; 0 other;
 Alignment Scores:
 Pred. No.: 8.65e-195 Length: 1528
 Score: 2414.50 Matches: 450
 Percent Similarity: 94.14% Conservative: 0
 Best Local Similarity: 94.14% Mismatches: 1
 Query Match: 98.79% Indels: 27
 DB: 20 Gaps: 1
 US-09-902-772-2 (1-451) x AAX26552 (1-1528)
 QY 1 MetLaserThriIeYsGluAlaLeuSerValValSerGluAspGlnSerLeuPheGlu 20
 DB 63 ATGGCAAGCACTATTAAAGAGCAATTCAGTGTGAGTGAMGACCAGTCTCTTCCAGTGA 122
 QY 21 CysAlaTyrGlySerProHisLeuAlaIysThrGluMetThrAlaSerSerSerGlu 40
 DB 123 TGTGCTTACGAGATGCCCTTGGCAAGACAAATGACAGCTCTCTTCCAGTGA 182
 QY 41 TTTGlyGlnThrSerLysMetSerProArgValProGlnGlnAspTyrLeuSerGlnPro 60
 DB 183 TATGGCAAAACATCAAGATGAGGCCGCCGGTCCACAGACGAGCTGTTATACACGCC 242
 QY 61 ProAlaIyrValThrIleYsMetGluCysAsnProAsnGlnValAsnGlySerArgAsn 80
 DB 243 CCGGCCGAGATTACCAATTAAGATGAGTGAACCCAAACAGCTTAATGGGTCAAGCAAT 302
 QY 81 SerProAspAspCysSerValAlaIalysGlyLysMetValSerSerSerAspAsnVal 100
 DB 303 TCACCTTATGACTGCACGCGGCAAAAGAGGAAATGTTTACACATTCGACAAAGT 362
 QY 101 GlyMetAsnTyrGlySerTyrMetGluGlnLysHisIleProProAsnMetThrThr 120
 DB 363 GGGATGAACATATGAGACCTACATGAGAGAGACATATTCGGCTCCAAATATGACAAAC 422
 QY 121 AsnGluArgValIleValProAlaAspProThrLeuTyrSerThrAspHisValArg 140

DB 423 AATGACGAAAGATTATTGTGCGACAGATCTCAGTTATGAGACACAGACCATGACGG 482
 QY 141 GlnTyrLeuGluTyrPalValIysGluTyrGlyLeuProAspValAspIleLeuPhe 160
 DB 483 CAGTGCTGAGGTGGCGACGTGAAGAGTATGCTTCCAGACGTGGACATCTTGTGTC 542
 QY 161 GlnAsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArgLeuThr 180
 DB 543 CAGAAACATTTGATGGGAAAGATTGTGTAATAATGACCAAAAGATGACTTCCACAGACTCAG 602
 QY 181 ProSerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGluArg----- 198
 DB 603 CCGAGCTATACGACGATATCTCTCTGACACCTACACTACCTCAGAGACTCTCTT 662
 QY 198 ----- 198
 DB 663 CCACATTTGACTTCAATGATGTTGATTAAGCCCTTACAAAACCTCCACGGTTAATGCAT 722
 QY 199 -----GlyAlaThrPheIlePheProAsnThrSerValTyrProGluAla 213
 DB 723 GCTAGAAACACAGGAGGAGCCACTTTTATTTTCCAAATACATCAGTTTACCAGAAAGCA 782
 QY 214 ThrGlnArgIleThrThrArgProAspLeuProTyrGluGlnAlaArgArgSerAlaTrp 233
 DB 783 ACGCAAAAGAAATMACAAAGCCAGATTACCTTATGAGCAAGCGAGAGATCAGCGTGG 842
 QY 234 ThrSerHisSerHisProThrGlnSerLysAlaThrGlnProSerSerSerThrValPro 253
 DB 843 ACGAGTACACGCCATCCACTCACTCACTCAAAAGCTAACCAACCATATCTTCAACAGTCCC 902
 QY 254 LysThrGluAspGlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSerSer 273
 DB 903 AAAACAGAAAGCCAGCGCTCCTAGTATGATCCTTATCAGATTCCTTGGACGACACGACGC 962
 QY 274 ArgLeuAlaAsnProGlySerGlyGlnIleGlnLeuTyrGlnPheLeuLeuLeu 293
 DB 963 CGTCTTCAAAATCCAGGAGAGTGGGCAATACAGTATGCGATTCCTCACTGAGACTTCTG 1022
 QY 294 SerAspSerSerAsnSerAsnCysIleThrTrpGluGlyThrAsnGlyGluPheLysMet 313
 DB 1023 TCGGACAGCTTCAACTCCAACTGCATGCATCCGGAAGGACAAATGGGGATTAAAGT 1082
 QY 314 ThrAspProAspGluValAlaArgArgTrpGlyGluArgLysSerLysProAsnMetAsn 333
 DB 1083 ACAGACCTGATGAAGTGGCTCGGCGTGGGAGAGAGAAACCAACCTAACATGTAAC 1142
 QY 334 TyrAspLysLeuSerArgAlaLeuArgTyrTyrTyrAspLysAsnIleMetThrLysVal 353
 DB 1143 TATGACAAACTCAGCCGCTGCACTTCGCTACTACTATGACAAAATATATTATGACTAAAGTT 1202
 QY 354 HisGlyLysArgTyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnPro 373
 DB 1203 CATGTAAAGCGATATGCCCTTCAAAATTTGATTTCCACGGAATCGCTCAGGCCCTTCAACCT 1262
 QY 374 HisProProGlnLeuSerMetTyrLysTyrProSerAspLeuProTyrMetSerSerTyr 393
 DB 1263 CACCCTCAGATATCATCATGATACAAATPACCAATCAGACCTCCCTTACATGATGTTCTAC 1322
 QY 394 HisAlaHisProGlnLysMetAsnPheValAlaProHisProProAlaLeuProValThr 413
 DB 1323 CATGCACACCCCGCAGAAAGATGAACCTTTGTAGCTCCATGCCCTGCTGTGCGCGTAAC 1382
 QY 414 SerSerSerPhePheAlaIalProAsnProTyrTyrAsnSerProThrGlyIleTyr 433
 DB 1383 TCATTCAGCTTTTGTGTCCTTAATCCATACCTGGAATTCACCACTGGAGGATCTAC 1442
 QY 434 ProAsnThrArgLeuProAlaAlaHisMetProSerHisLeuGlyIleThrTyr 451
 DB 1443 CCGAATACCGAGCGTGCAGTGCATATGCTTCCATCTTGGCACCTACTAC 1496
 RESULT 3
 AA050644

ID	AA050644 standard; cDNA; 2938 BP.
XX	AA050644;
AC	26-MAY-1994 (first entry)
DT	Human Hum-Fil-1 gene clone BM025.
XX	
DE	chromosomal translocation; chimeric; chimeric; Ewing sarcoma;
XX	Ews gene; malignant melanoma; hum-fil-1;
KW	primitive peripheral neuroectodermal tumour; human chromosome 11;
KW	human chromosome 22; ss.
XX	
OS	Homo sapiens.
XX	
Key	Location/Qualifiers
EH	143..1501
FT	/*tag= a
FT	/product= HUM-FIL-1
FT	2908..2913
FT	polya_signal
FT	/*tag= b
XX	
XX	W09323549-A.
XX	
PD	25-NOV-1993.
XX	
PF	19-MAY-1993; 93WO-FR00494.
XX	
PR	20-MAY-1992; 92FR-0006123.
XX	
PA	(CNRS) CNRS CENT NAT RECH SCT.
XX	
PI	Aurias A, Delattre O, Desmaza C, Melot T, Peter M;
PI	Plougastel B, Thomas G, Zucman J;
XX	
DR	WPI: 1993-386580/48.
XX	P-PSDB; AAR44556.
XX	
XX	New nucleic acid of EWS gene and its hybrid(s) - contg. gene
PT	sequence involved in chromosomal trans-location, also derived
PT	mRNA, probes, fusion proteins etc., for diagnosis and treatment
PT	of Ewing sarcoma and melanoma
XX	
PS	Disclosure; Fig 7; 123pp; French.
XX	
CC	The probe 11R1 was used to screen a human marrow cDNA library
CC	(Clontech cat. # HL1058). The clone BM025 was identified and
CC	sequenced. It represents the entire coding region together with
CC	5'- and 3'-UTRs of the Hum-Fil-1 gene.
XX	
SQ	Sequence 2938 BP; 847 A; 692 C; 675 G; 724 T; 0 other;
XX	
Alignment Scores:	
Pred. No.:	1, 51e-127 Length: 2938
Score:	1624.50 Matches: 306
Percent Similarity:	79.43% Conservative: 57
Best Local Similarity:	66.96% Mismatches: 83
Query Match:	66.47% Indels: 11
DB:	14 Gaps: 7
US-09-902-772-2 (1-451) x AA050644 (1-2938)	
QY	1 Meta1aserTrilleytysgluAlaIeuserYalValsergluaspInserleupheglu 20
Db	143 ATGACGGGACATTATTAAAGAGGCTCTGTGGTGTGAGCAGCAGCTCCCTTTTGAC 2022
QY	21 CysAlaTyrGly---SerProHisLeuAlaLysThrGluMetThrAlaSerSerSer 39
Db	203 TCAGAGTCAGGAGGCGAGCCACATCTGCCCAAGCCAGCAATGACTGCTCGGGGAGTCT 2622
QY	40 GluTyrGlyGlnThrSerLysMetSerProArgValProGlnGlnAspTrpLeuSerGln 59
Db	263 GACATACGGGACGCCACACATCAACCCCTCCACACAGCAGGAGGTGATCAATACAG 3222

QY	60	ProProAlaAargValThrIleIleIysMeGluCysAsnProAsnGlnValAsnGlySerArg	79
Db	323	--CCAGTGAAGGAGTCAACGTCMAACGGGAGATAT--GACCACCAATGAAATGATCCAGG	373
QY	80	AsnSerProAspAspCysSerSerValAlaIleGlyGlyGlyMetValSerSerSerAspAsn	99
Db	374	GAGTCTCCGGTGGACTGCGAGCGTTAGCAAAATGACAGCAAGCTGTGGCGGAGGCCAGTCC	433
QY	100	ValGlyMetAsnIleYrgIleSerTyrMetGluGluIleHis--IleProProProAsnMet	118
Db	434	AACCCCAAGAACCTCAACACAGTATATGAGAGAGAAAGATGGCCCCCTCTCCCAACAAG	493
QY	119	ThrThrAsnGluAArgArgValIleValProAlaAspProThrLeuTyrSerThrAspHis	138
Db	494	ACACACCAACGAGAGAGAGAGTATGTCTCCCGCAGACCCCACTGTGGACACAGAGAGCT	553
QY	139	ValAArgIleTyrIleuGluTyrProAlaValAlaIleGlyGluTyrGlyLeuProAspValAspIleLeu	158
Db	554	GTGAGAGCAATGGCTGGAGTGGGGCCATAAAGAGATATAGCTTGATGGAGATCGACACATCC	613
QY	159	LeuPheGlnAsnIleAspArgIleuGluIleuGlyMetThrIleAspAspAspPheGlnArg	178
Db	614	TTTTTCCGAACATGATGATGGCAAGAACTGTGTAAATGTAAACAAGAGAGACTTCCCTCCC	673
QY	179	LeuThrProSerTyrAsnAlaAspIleLeuLeuSerHisIleuHisIleuTyrLeuArgIleArg	198
Db	674	GCCACACCCCTCTACAAACAGGAAGTCTGTGTCAACCTCACTGATACCTCAGGAA--	730
QY	199	GlyAlaThrPheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgIleThr	218
Db	731	--AGTTACACGTGGGGCTTAATATCAACCTCCCAACCGCAACCAATCTCCAGATTTGAGT	787
QY	219	ThrArgProAspLeuProTyrGluGlnAlaArgArgSerAlaThrPheThrSerHisSerHis	238
Db	788	GTCAAAGAAAGACCTTCTTATGACTCAGTACAGAGAGAGACATGGGGCAATATACATGAT	847
QY	239	ProThrGlnSerIleValIleArgin--ProSerSerSerThrValProIleProIleThrGluAsp	257
Db	848	TCTGGCCCAACAAATACTCTCCCTCCGAGGGGGCAACAAACGATCAGTAAGAATACAGAG	907
QY	258	GlnArgProGluIleuAspProTyrGlnIleIleuGlyProThrSerSerArgIleuAlaAsn	277
Db	908	CACAGGCCCCACGACAGATCCGTATCAATCTGTGGCCGACGACGTGCGCTTACGCAAC	967
QY	278	ProGlySerGlyGlnIleGlnLeuTyrGlnPheLeuLeuGluIleuLeuSerAspSerSer	297
Db	968	CCTGGAACGGGCGAGATCCAGCTGTGCATATCTCCTGAGCTGCTTCCGACAGCGCC	1027
QY	298	AsnSerAsnGlyIleThrTyrPheGluGlyThrAsnGlyGluPheIleuSerThrAspProAsp	317
Db	1028	AACGCGACGCTGTATCACTGTGGAGGGGACCAACGGGAGATCTAAATATGACGAGACCCGAT	1087
QY	318	GluValAlaAArgArgTyrPheGlyGluArgIleSerIleProAsnMetAsnIleTyrAspIleu	337
Db	1088	GAGGGGCGACAGCGCTGTGGGGGAGCGGAAAGAACAACCCCACTGAATATACGACAACGCTG	1144
QY	338	SerArgAlaIleuArgIleTyrTyrAspIleAsnIleMetThrIleValHisIleGlyIleArg	357
Db	1148	AGCGGGCGCCCTCTTATCTATGATGATAAACATTTATGACCAAACTGACAGGCAAAAGA	1207
QY	358	TyrAlaIleTyrIleAspPheAspPheHisGlyIleAlaGlnAlaIleuGlnProHisProGlu	377
Db	1208	TATCTTTACAAATTTGACTTCCACGGCATTTGCCAGGCTCTGAGGCACATCTCCAGCGAG	1267
QY	378	SerSerMetTyrIleTyrProSerAspLeuProTyrMetSerSerTyrHisIleHisPro	397
Db	1268	TGCGTCATGTACAGATACCTTCTGCACATCTCTACATGCTTCCATACATCCCAACGACG	1327
QY	398	GlnIleuSerAsnPheValAlaProHisProProAlaLeuProValIleThrSerSerSerPhe	417
Db	1328	CAGAGGTGAACCTTTGTCCCTCCCAACATCTCTCCATGCTGTGTCACTTCTCTCCAGCTTC	1387


```

OY 239 ProthrglnSerLysAlaThrGln---ProSerSerSerThrValProLysThrGlnAsp 257
DB 878 TCTGGCCCTCAACAAAAGTCTCCCTTGGAGGGGCACAAACGATCATGAATACAGAG 937
OY 256 GlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSerSerArgLeuAlaAsn 277
DB 938 CAACGGCCCGCCAGATCGTATCATGATCCGGCCCGACAGCATGCTGCTAGCCAAAC 997
OY 278 ProGlySerGlyGlnIleGlnLeuTrpGlnPheLeuLeuGlnLeuLeuSerAspSer 297
DB 998 CCTGAGAGGGGACATCCAGCTGTGGCAATCTCTCTGAGCTGCTCTCCAGACGGCC 1057
OY 298 AsnSerAsnCysIleThrTrpGlnGlyThrAsnGlyGlnPheLysMetThrAspProAsp 317
DB 1058 AACGCCAGCTGATCATCCTGGAGGAGGACCAACGGGAGCTTCAAAATGACGACCCCGAT 1117
OY 318 GluValAlaAlaArgArgTrpGlyLysArgLysSerLysProAsnMetAsnTyrAspLysLeu 337
DB 1118 GAGGTGGCCAGCGCTGGGGCGAGCGAAGCAAGCCCAACATGAAATTACAGCAAGCTG 1177
OY 338 SerArgAlaLeuArgTyrTyrTyrAspLysAsnIleMetThrLysValHisGlyLysArg 357
DB 1178 AACCGGGCCCTCGCTTATCTATGATAAAACATTATATACCAAGTGCACGGCAAAAGA 1237
OY 358 TyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnProHisProGln 377
DB 1238 TATGCTTCAAAATTTGACTTCCACGCGCATTCGCCAGGCTCTGACGACCATCCGACCGAG 1297
OY 378 SerSerMetTyrLysTyrProSerAspLeuProTyrMetSerSerThrLysAlaHisPro 397
DB 1298 TGTGTCATATACAGTACCTCTCTGACATCTCTTACATGCTCTCTCTACCATGCCCCACAG 1357
OY 398 GlnLysMetAsnPheValAlaProHisProProAlaLeuProValThrSerSerPhe 417
DB 1358 CAAAGAGTCAACTTGTCTCCCTCCCATCCATCCATCCATCCATCCATCCATCCATCC 1417
OY 418 PheAlaAlaProAsnProTyrTyrPheAsnSerProThrGlyGlyIleTyrProAsn----- 435
DB 1418 TTTGGACGGCATCAACAAATACAGGACCTCCCGACGGGGGAGATCAACCCCAACCCCAAC 1477
OY 436 ---ThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr 451
DB 1478 GTCCCCCGGCATCTACACCCACGCTGCTTACACTTACAGGAGCTACTAC 1528

```

```

XX WPI: 2001-266308/27.
DR P-PSDB; AAB90792.
XX
XX DNA sequences, proteins encoded by them and antibodies against them
PT useful in diagnosis and treatment of vascular disease caused by
PT arteriosclerosis -
XX
PS Claim 20; Page 462-466; 678bp; Japanese.
XX
CC The present invention provides the protein and coding sequences of a
CC number of human shear stress response proteins. These are useful in the
CC diagnosis, treatment and screening of vascular diseases caused by
CC arteriosclerosis, including heart failure, post-PTCA restenosis and
CC hypertension.
XX
SQ Sequence 2954 BP; 846 A; 698 C; 683 G; 727 T; 0 other:

```

Alignment Scores:

Pred. No.:	2,196-125	Length:	2954
Score:	1599.00	Matches:	304
Percent Similarity:	78.99%	Conservative:	57
Best Local Similarity:	66.52%	Mismatches:	84
Query Match:	65.43%	Indels:	12
DB:	22	Gaps:	8

US-09-902-772-2 (1-451) x AAB02915 (1-2954)

```

OY 1 MetAlaSerThrIleLysGlnAlaLeuSerValIleSerGluAspGlnSerLeuPheGlu 20
DB 173 ATGGAGGGAGCACTATTAAAGAGGCTCTGTGGTGGAGCGAGACAGTCCCTCTTGTGAC 232
OY 21 CysAlaTyrGly---SerProHisLeuAlaLysThrGlnMetThrAlaSerSerSer 39
DB 233 TCAGCGTACGAGGAGCGGACCCATCTCCCAAGGCGGACATGATGCTCGGGAGTCTCT 292
OY 40 GluTyrGlyGlnThrSerLysMetSerProArgValProGlnGlnAspTrpLeuSerGln 59
DB 293 GACTACGGGACCGCCACAAAGATCAACCCCTCCACACACAGAGAGTGCATCATGAG 352
OY 60 ProProAlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArg 79
DB 353 ---CGAGTGGGGTCAACCTCAAGCGGGAGTAT-----GACCACATGATGATCCAGG 403
OY 80 AsnSerProAspAspCysSerValAlaLysGlyLysMetValSerSerSerAspAsn 99
DB 404 GAGTCTCCGCTGACGTGACGCGCTTAGCAAAATGCACCAAGCTGTGGCGGAGCGACATCC 463
OY 100 ValGlyMetAsnTyrGlySerTyrMetGluGlySHis---IleProProProAsnMet 118
DB 464 AACCCCATGAACTACACAGCTATATGACGAGAAATGAGCCCTCTCCCAACATG 523
OY 119 ThrThrAsnGluArgArgValIleValProAlaAspProThrLeuTrpSerThrAspHis 138
DB 524 ACCACCAACGAGAGAGATCATGCTGCCCGACGACCCACATGCTGGACACAGAGAT 583
OY 139 ValArgGlnTrpLeuGlnTrpAlaValLysGluTyrGlyLeuProAspValIleLeu 158
DB 584 GTGAGGCAATGCTGTGAGTGGCCATAAAGAGTACAGCTTGTATGAGATGCACACATCC 643
OY 159 LeuPheGlnAsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArg 178
DB 644 TTTTTCACAGAACATGATGAGCAAGAGAACTGTCTAAATGAAACAGAGAGACTTCTCCGC 703
OY 179 LeuThrProSerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGlnArg 198
DB 704 GCCACACACCTTACACACGGAAGTGTGTGTACACCTGATGTTACTTCAGGAA--- 760
OY 199 GlyAlaThrPheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgTyrIleThr 218
DB 761 ---AGTTCACGTCTGGCCCTATATACACCTCCACACCGACCAATCTCTCAGATTTGAGT 817
OY 219 ThrArgProAspLeuProTyrGlnIleAlaArgArgSerAlaTyrThrSerHisSerHis 238

```

```

Db 818 GTCAAGAGACCCCTTCTTATGACTCAGTCAAGAGAGAGCTGGGGCATATACATGAT 877
Qy 239 ProThrgInSerLysAlaThrgIn---ProSerSerThrgValProLysThrgInuAsp 257
Db 878 TCTGGCCCTCAACAAAGTCCCTCCCTGGAGGGGACAAAGATCAGTACAGATACAGAG 937
Qy 258 GluArgProGlnLeuAspProTyrgInLeuGlnProThrgSerArgLeuAlaAsn 277
Db 938 CAAGCGCCCGCCAGATCCGATCAGATCCGAGCCCGACACAGACAGCCGCTAGCCAAAC 997
Qy 278 ProGlySerGlnInLeuGlnLeuThrgInPheLeuLeuGlnLeuLeuSerAspSerSer 297
Db 998 CCTGGAAGCGGCGAGATCCAGCTGTGGCAATCTCTCTGGAGCTGCTCCGACAGCGCC 1057
Qy 298 AsnSerAsnCySLeThrgTrpGluGlyThrasnGlyGluPheLysMetThrgAspProasp 317
Db 1058 AAGCCAGCTGTATCAGCTGGAGGGAGCCACAGGGAGTTCAAAATAGACGAGACCCGAT 1117
Qy 318 GluValAlaArgArgTrpGlyGluArgLysSerLysProAsnMetAsnTyrgAspLysLeu 337
Db 1118 GAGTGGCGCAGCGCTGGGGGAGCGGAAAGCAAGCCCAATGAAATTAGACAGCTG 1177
Qy 338 SerArgAlaLeuArgTyrgTyrgTyrgAspLysAsnIleMetThrgValHisGlyLysArg 357
Db 1178 AGCGGGCCCTCCGTTATGATATGATATGATATGATATGATATGATATGATATGATATG 1237
Qy 358 TyrgAlaTyrgLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnProHisProProGlu 377
Db 1238 TATGCTTACAAATTTGACTTCCAGCGCATTCGCCAGCTCGACGCCATCCAGCCAGAG 1297
Qy 378 SerSerMetTyrgLysTyrgProSerAspLeuProTyrgMetSerSerTyrgHisAlaHisPro 397
Db 1298 TCGTCCATGTGCAAGTACCCCTTGACATCTCTACATCTCTACATCTCTACATCTCTAC 1357
Qy 398 GlnLysMetAsnPheValAlaProHisProProAlaLeuProHisProHisSerSerPhe 417
Db 1358 CAGAAAGGTGAAGCTTGTCCCTCCCATCCATCCATCCATCCATCCATCCATCCATCCAT 1417
Qy 418 PheAlaAlaProAsnProTyrgTrpAsnSerProThrgGlyIleTyrgProAsn----- 435
Db 1418 TTGGAGCCCGATCACAATATGAGACCTCC---ACGGGGGGAATCTACCCCAACCCCAAC 1474
Qy 436 ---ThrgArgLeuProAlaAlaHisMetProSerHisLeuGlyThrgTyrg 451
Db 1475 GTCCCGCCGATCCTTAACACCCAGCTGCTTACACATTAGAGCAGCTACTAC 1525

RESULT 6
AA050662
ID AA050662 standard; DNA; 1890 BP.
XX
AC AA050662;
XX
DT 26-MAY-1994 (first entry)
XX
DE Human Hum-Fil-1 gene (genomic DNA).
XX
KW chromosomal translocation; chimeric; chimaeric; Ewing sarcoma;
KW Ews gene; malignant melanoma; hum-fil-1;
KW primitive peripheral neuroectodermal tumour; human chromosome 11;
KW human chromosome 22; ss.
XX
OS Homo sapiens.
XX
FH Key location/Qualifiers
FT 1..39
FT exon /*tag= a
FT /*number= 1
FT intron /codon_start= 22..24
FT /*tag= b
FT /*number= 1
FT exon 100..311

```

```

FT /*tag= c
FT /number= 2
FT intron 312..371
FT /*tag= d
FT /number= 2
FT exon 372..526
FT /*tag= e
FT /number= 3
FT intron 527..586
FT /*tag= f
FT /number= 3
FT exon 587..790
FT /*tag= g
FT /number= 4
FT intron 791..850
FT /*tag= h
FT /number= 4
FT exon 851..916
FT /*tag= i
FT /number= 5
FT intron 917..976
FT /*tag= j
FT /number= 5
FT exon 977..1042
FT /*tag= k
FT /number= 6
FT intron 1043..1102
FT /*tag= l
FT /number= 6
FT exon 1103..1162
FT /*tag= m
FT /number= 7
FT intron 1163..1222
FT /*tag= n
FT /number= 7
FT exon 1223..1270
FT /*tag= o
FT /number= 8
FT intron 1271..1330
FT /*tag= p
FT /number= 8
FT exon 1331..1890
FT /*tag= q
FT /number= 9
FT
FN WO9323549-A.
XX
PD 25-NOV-1993.
XX
PF 19-MAY-1993; 93MO-FR00494.
XX
PR 20-MAY-1992; 92FR-0006123.
XX
PA (CNRS ) CNRS CENT NAT RECH SCI.
XX
PI Aurias A, Delattre O, Desmazes C, Melot T, Peter M;
XX
PI Plougastel B, Thomas G, Zucman J;
XX
DR WPI: 1993-386580/48.
XX
DR P-PSDB; AAR44556.
XX
PT New nucleic acid of EWS gene and its hybrid(s) - contg. gene
PT sequence involved in chromosomal trans-location, also derived
PT mRNA, probes, fusion proteins etc., for diagnosis and treatment
PT of Ewing sarcoma and melanoma
XX
PS Disclosure; Fig 7 and Fig 13; 123pp; French.
XX
CC The intron-exon junctions of the human Ews gene and the Hum-Fil-1
CC gene have been sequenced (see AA050646 and AA050662, respectively).
CC The different fusion products which could be formed by fusing
CC exons from the two genes, as happens after specific chromosomal
CC translocations, can be predicted (see AA050671-Q50678). The sequences

```



```

DB 114 TCCAGTTTGTGCTGCCCCAACCCATCTGGAATTCACCAACTGGGGGTATATACCC 55
QY 435 AsnThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr 451
DB 54 AACACTAGGCTCCGCCACACGACCATATGCTTCATCTGGGCACTTACTAC 4

RESULT 10
AAK14426/c
ID AAK14426 standard; DNA: 567 BP.
XX
AC AAK14426;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human brain expressed single exon probe seq ID NO: 14417.
XX
KW Human: brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200157275-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00667.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483446/52.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
PT brains -
XX
PS Example 4; SEQ ID NO: 14417; 650bp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is one of the probes of the
CC invention.
XX
SQ Sequence 567 BP; 114 A; 124 C; 200 G; 129 T; 0 other;

Alignment Scores:
Pred. No.: 2,91e-71 Length: 567
Score: 948.00 Matches: 169
Percent Similarity: 97.74% Conservative: 4
Best Local Similarity: 95.48% Mismatches: 4
Query Match: 38.79% Indels: 0
DB: 22 Gaps: 0

US-09-902-772-2 (1-451) x AAK14426 (1-567)
QY 275 LeuAlaAsnProGlySerGlyGlnIleGlnLeuTrpGlnPheLeuLeuGlnLeuSer 294
DB 534 CTGACCTCATATAGGACAGTGCAGATCCACATTGGCAGTCTCTCCGAGAGCTCCTG 475
QY 295 AspSerSerAsnSerAsnCysIleThrTTPGLUGlyThrAsnGlyIuphelyMetThr 314

```

```

DB 474 GACAGCTCCCAACTCCACTGCATCACCCTGGGAAGCAACCGGGGATTCATCATGACG 415
QY 315 AspProAspGluValAlaAlaArgTyrGlyIuArgLysSerLysProAsnMetAsnTyr 334
DB 414 GATCCCAACGACGAGTGGCCCGGCGCTGGGAGAGCGGAGACCAACCCATCATGACACTAC 355
QY 335 AspLysLeuSerArgAlaLeuArgTyrTyrAspLysAsnIleMetThrLysValHis 354
DB 354 GATAGCTCAGCCGCGCCCTCCGTACTACTATGACAAGAACATCATGACCAAGTCCAT 295
QY 355 GlyLysArgTyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnProHis 374
DB 294 GGGAAAGCGCTACGCGCTACAGTTGACTTCACCGGATCGCCGAGCCCTCCACACCCAC 235
QY 375 ProProGlySerSerMetTyrLysTyrProSerAspLeuProTyrMetSerSerTyrHis 394
DB 234 CCCCCGAGTCATCTCTGTACAGTACCCCTCAGACCTCCCGTACATGGGCTCTATCAC 175
QY 395 AlaHisProGlnLysMetAsnPheValAlaProHisProProAlaLeuProValThrSer 414
DB 174 GCCCACCACAGAGATGAATCTTGTGGCGCCACCCCTCCAGCCCTCCCGTACATCT 115
QY 415 SerSerPhePheAlaAlaProAsnProTyrTTPAsnSerProThrGlyIleTyrPro 434
DB 114 TCCAGTTTGTGCTGCCCCAACCCATATGCTTCATCTGGGCACTTACTAC 55
QY 435 AsnThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr 451
DB 54 AACACTAGGCTCCGCCACACGACCATATGCTTCATCTGGGCACTTACTAC 4

RESULT 11
AAK40160/c
ID AAK40160 standard; DNA: 567 BP.
XX
AC AAK40160;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human bone marrow expressed single exon probe SEQ ID NO: 14717.
XX
KW Human: bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukemia; lymphoma; myeloma; ss.
XX
OS Homo sapiens.
XX
PN WO200157276-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00668.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488900/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human bone marrow -
XX
PS Example 4; SEQ ID NO: 14717; 658bp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human

```

CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukemia and myeloma. The present sequence is one of
CC the probes of the invention.

SO Sequence 567 BP; 114 A; 124 C; 200 G; 129 T; 0 other;

Alignment Scores:

Pred. No.:	2,91e-71	Length:	567
Score:	948.00	Matches:	169
Percent Similarity:	97.74%	Conservative:	4
Best Local Similarity:	95.48%	Mismatches:	4
Query Match:	38.79%	Indels:	0
DB:	22	Gaps:	0

US-09-902-772-2 (1-451) x AAK40160 (1-567)

```

QY 275 LeuAlaAsnProGlySerGlyGlnIleGlnLeuTrpGlnPheLeuLeuGlnLeuLeuSer 294
    ||| ::: |||||
DB 534 CTGACCTCATTTAGGCGAGTGGCCAGATCCAGCTTTGGCAGTCTCTGGAAGCTCTGTG 475

QY 295 AspSerSerAsnSerAsnGlyIleThrTrpGluGlyThrAsnGlyLuphelysMetThr 314
    |||||
DB 474 GACAGCTCAACTCCAGGTCGATCAGCTGGGAAGGACACCAAGGGAGTTCAAGATGAGC 415

QY 315 AspProAspGluValAlaIaArgArgTrpGlyGluArgLysSerLysProAsnMetAsnTyr 334
    |||||
DB 414 GATCCCGAGAGGTGGCCCGGCTGGGAGAGCGGAGCAAGCAACCAATGAACTAC 355

QY 335 AspLysLeuSerArgAlaLeuArgTrpGlyTyrTyrAspLysAsnIleMetThrLysValHis 354
    |||||
DB 354 GATTAAGCTACGCGCGCCCTCCCTTACTACTATGACAAAGACATGACCAAGGTCCAT 295

QY 355 GlyLysArgTyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnProHis 374
    |||||
DB 294 GGGAAAGCGCTAGCGCTACAAAGTTGCGACTTCACAGGGATGCCCGCCAGCCCGCCAC 235

QY 375 ProProGlnSerSerMetLysTyrLysTyrProSerAspLeuProLysMetSerTyrHis 394
    |||||
DB 234 CCCCCGAGTCATCTGTGACAAAGTACCCCTCAGACCTCCGTCATGCGGCTCTATAC 175

QY 395 AlaHisProGlnLysMetAsnPheValAlaProHisProProAlaLeuProValThrSer 414
    |||||
DB 174 GCCCACCCACAGAAAGATGAATTTGGGCGCCCGCCAGCTCCAGCCCTCCCGGACATCT 115

QY 415 SerSerPhePheAlaAlaProAsnProTyrTyrPAsnSerProThrGlyLysIleTyrPro 434
    |||||
DB 114 TCCAGTTTTTTGGTCCCGCCAAACCATGAGTGAATTCACCAACGTGGGGGTATATACCC 55

QY 435 AsnThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr 451
    |||||
DB 54 AACACTAGGCTCCCGCCACGACCATATGCTTCATCTGCGGCGCTTACTAC 4

RESULT 12
AAI20932/C
ID AAI20932 standard; DNA; 567 BP.
XX
XX AAI20932;
XX
XX 12-OCT-2001 (first entry)
DE Probe #10865 for gene expression analysis in human cervical cell sample.
XX
XX Probe: human; microarray; gene expression; cervical epithelial cell;
XX
XX cervical cancer; ss.
XX
XX Homo sapiens.
XX
XX WO200157278-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00670.
XX
XX PF

```

```

XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488901/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human cervical epithelial cells -
XX
XX Claim 25; SEQ ID No 10865; 487pp; English.
XX
XX The present invention relates to human single exon nucleic acid probes
CC (SENP). The present sequence is one such probe. The SENPs are derived
CC from human HeLa cells. The SENPs can be used to produce a single exon
CC microarray, which can be used for measuring human gene expression in a
CC sample derived from human cervical epithelial cells. By measuring gene
CC expression, the probes are therefore useful in grading and/or staging
CC of diseases of the cervix, notably cervical cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 567 BP; 114 A; 124 C; 200 G; 129 T; 0 other;
SO
XX
XX Alignment Scores:
XX Pred. No.: 2,91e-71 Length: 567
XX Score: 948.00 Matches: 169
XX Percent Similarity: 97.74% Conservative: 4
XX Best Local Similarity: 95.48% Mismatches: 4
XX Query Match: 38.79% Indels: 0
XX DB: 22 Gaps: 0
XX
XX US-09-902-772-2 (1-451) x AAI20932 (1-567)
QY 275 LeuAlaAsnProGlySerGlyGlnIleGlnLeuTrpGlnPheLeuLeuGlnLeuLeuSer 294
    ||| ::: |||||
DB 534 CTGACCTCATTTAGGCGAGTGGCCAGATCCAGCTTTGGCAGTCTCTGGAAGCTCTGTG 475

QY 295 AspSerSerAsnSerAsnGlyIleThrTrpGluGlyThrAsnGlyLuphelysMetThr 314
    |||||
DB 474 GACAGCTCAACTCCAGGTCGATCAGCTGGGAAGGACACCAAGGGAGTTCAAGATGAGC 415

QY 315 AspProAspGluValAlaIaArgArgTrpGlyGluArgLysSerLysProAsnMetAsnTyr 334
    |||||
DB 414 GATCCCGAGAGGTGGCCCGGCTGGGAGAGCGGAAAGCAACCAATGAACTAC 355

QY 335 AspLysLeuSerArgAlaLeuArgTrpGlyTyrTyrAspLysAsnIleMetThrLysValHis 354
    |||||
DB 354 GATTAAGCTACGCGCGCCCTCCCTTACTACTATGACAAAGACATGACCAAGGTCCAT 295

QY 355 GlyLysArgTyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnProHis 374
    |||||
DB 294 GGGAAAGCGCTAGCGCTACAAAGTTGCGACTTCACAGGGATGCCCGCCAGCCCGCCAC 235

QY 375 ProProGlnSerSerMetLysTyrLysTyrProSerAspLeuProLysMetSerTyrHis 394
    |||||
DB 234 CCCCCGAGTCATCTGTGACAAAGTACCCCTCAGACCTCCGTCATGCGGCTCTATAC 175

QY 395 AlaHisProGlnLysMetAsnPheValAlaProHisProProAlaLeuProValThrSer 414
    |||||
DB 174 GCCCACCCACAGAAAGATGAATTTGGGCGCCCGCCAGCTCCCGCGGACATCT 115

QY 415 SerSerPhePheAlaAlaProAsnProTyrTyrPAsnSerProThrGlyLysIleTyrPro 434
    |||||

```

```

Db 114 TCCAGTTTTTTCGTCGCCCAACCCATCTGGAATTCACCACTGGGGGTATATACCCC 55
QY 435 AsnThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyr 451
    |||||
Db 54 AACACTAGGCGTCCCGACACGACATATGCTTTCATCTGGGCACTTACTAC 4
RESULT 13
AAI46176/c
ID AAI46176 standard; DNA; 567 BP.
XX
AC AAI46176;
XX
DN 17-OCT-2001 (first entry)
XX
DE Probe #14862 used to measure gene expression in human placenta sample.
XX
KM Probe: microarray; human; placenta; antenatal diagnosis;
XX genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN W0200157272-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00663.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488897/53.
XX
XX
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta.
XX
XX
XX Claim 25; SEQ ID NO 14862; 654pp; English.
XX
CC The present invention relates to single exon nucleic acid probes (SENP).
CC CC Producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX
SQ Sequence 567 BP; 114 A; 124 C; 200 G; 129 T; 0 other;
Alignment Scores:
Pred. No.: 2,91e-71 Length: 567
Score: 948.00 Matches: 169
Percent Similarity: 97.74% Conservative: 4
Best Local Similarity: 95.48% Mismatches: 4
Query Match: 38.79% Indels: 0
DB: 22 Gaps: 0
US-09-902-772-2 (1-451) x AAI46176 (1-567)
QY 275 IeuAlaAsnProGlySerGlyGlnIleGlnLeuTrpGlnPheLeuLeuGlnLeuSer 294
    |||
Db 534 CTGACCTCATTTAGGCACTGGCCAGATCTCAGCTTTGGCAGTTCTCTCGGAGCTCTGTCG 475
QY 295 AspSerSerAsnSerAsnCysIleThrTrpGlnGlyThrAsnGlyGlnPheLeuMetThr 314
    |||||
Db 474 GACAGCTCAACTCAGCTCAGCTCAGCTGGAAGGACCAAGGGGAGTTCAAGATGACG 415
QY 315 AspProAspGlnValAlaAlaArgAlaGlyTrpGlyGlnArgGlySerIysProAsnMetAsnTyr 334

```

```

Db 414 GATCCCGACGAGGTGGCCCGCGCTGGGAGAGCGGAGACCAACCCATCATACTAC 355
QY 335 AspLysLeuSerAArgAlaLeuArgTyrTyrTrpAspLysAsnIleMetThrLysValHis 354
    |||||
Db 354 GATAGCTCAGCCCGCCCTCCGTTACTACTATGACACAGACATCATGACCAAGGTCAT 295
QY 355 GlyLysArgTyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnProHis 374
    |||||
Db 294 GGGAAGGCTACGGCTCAAGTTGACTTCACAGGAGATCGCCAGGCGCTCCACGCCAC 235
QY 375 ProProGlnSerSerMetTyrLysTyrProSerAspLeuProTyrMetSerSerTyrHis 394
    |||||
Db 234 CCCCCGAGTCATCTGTACAACTACCCCTCAACCTCCCGTACATGCGCTCTATAC 175
QY 395 AlaHisProGlnLysMetAlaPheValAlaProHisProProAlaLeuProValThrSer 414
    |||||
Db 174 GCCCACCACAGAAAGATGAACTTGTGGCGCCACCCCTCCAGCCCTCCGTACACT 115
QY 415 SerSerPhePheAlaAlaProAsnProTyrTrpAsnSerProThrGlyGlyIleTyrPro 434
    |||||
Db 114 TCCAGTTTTTTCGTCGCCCAACCCATCTGGAATTCACCACTGGGGGTATATACCCC 55
QY 435 AsnThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyr 451
    |||||
Db 54 AACACTAGGCGTCCCGACACGACATATGCTTTCATCTGGGCACTTACTAC 4
RESULT 14
AAI06643/c
ID AAI06643 standard; DNA; 567 BP.
XX
XX AAI06643;
XX
XX
XX 09-OCT-2001 (first entry)
XX
DE Probe #6634 used to measure gene expression in human breast sample.
XX
XX Probe: human; breast disease; breast cancer; development disorder; ss;
XX inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
XX Homo sapiens.
XX
XX
XX W0200157270-A2.
XX
XX
XX 09-AUG-2001.
XX
XX
XX 29-JAN-2001; 2001WO-US00661.
XX
XX
XX 04-FEB-2000; 2000US-0180312.
XX
XX 26-MAY-2000; 2000US-0207456.
XX
XX 30-JUN-2000; 2000US-0608408.
XX
XX 03-AUG-2000; 2000US-0632366.
XX
XX 21-SEP-2000; 2000US-0234687.
XX
XX 27-SEP-2000; 2000US-0236359.
XX
XX 04-OCT-2000; 2000GB-0024263.
XX
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-476286/51.
XX
XX
XX Novel single exon nucleic acid probe used to measuring gene expression
XX in a human breast.
XX
XX Claim 25; SEQ ID NO 6634; 322pp; English.
XX
CC The present invention relates to novel single exon nucleic acid probes.
CC CC The present sequence is one such probe. The probes are useful for
CC measuring human gene expression in a human breast sample, where the probe
CC hybridizes at high stringency to a nucleic acid expressed in the human
CC breast. The probes are useful for predicting, diagnosing, grading,
CC staging, monitoring and prognosis diseases of the human breast,

```

CC particularly those diseases with polygenic aetiology. The diseases
 CC include: breast cancer, disorders of development, inflammatory diseases
 CC of the breast, fibrocystic changes, proliferative breast disease and
 CC non-carcinoma tumours.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 567 BP; 114 A; 124 C; 200 G; 129 T; 0 other;

Alignment Scores:

Pred. No.:	2,91e-71	Length:	567
Score:	948.00	Matches:	169
Percent Similarity:	97.74%	Conservative:	4
Best Local Similarity:	95.48%	Mismatches:	4
Query Match:	38.79%	Indels:	0
DB:	22	Gaps:	0

US-09-902-772-2 (1-451) x AA106643 (1-567)

QY 275 LeuAlaAsnProGlySerGlyGlnIleGlnLeuTrpGlnPheLeuLeuGlnLeuSer 294
 DB 534 CTGACCTCATAGGAGTGGCCAGATCCAGCTTGGCAGTCTCTGGAGCTCTGTCG 475
 QY 295 AapSerSerAsnSerAsnGlySerIleThrTrpGlnGlyThrAsnGlyLeuPheLeuMetThr 314
 DB 474 GACAGTCCAACTCCAGCTGATCACCCTGGGAGGACCAACGGGAGTTCAGATGACG 415
 QY 315 AapProAspGlnValAlaIleArgTrpGlyGluArgGlySerLeuProAsnMetAsnTrp 334
 DB 414 GATCCCGAGAGAGTGGCCGCCCTGGGAGAGCCGGAAGACCAACCATGAACTAC 355
 QY 335 AapPLeuSerArgAlaLeuArgTrpTrpAspPLeuAsnIleMetThrLeuValHis 354
 DB 354 GATAACTCAGCGCGCCCTCCCTACTACTATGACAAAGACATCATGACCAAGTCCAT 295
 QY 355 GlyLeuArgTrpAlaLeuTrpPheAspPheHisGlyIleAlaGlnAlaLeuGlnProHis 374
 DB 294 GGGAGCGGTACCGCTTACAAAGTTCGACTTCACGGGATGCGCCAGCCCTCCAGCCCCAC 235
 QY 375 ProProGlnSerSerMetTrpTrpProSerAspLeuProTrpMetSerSerTrpHis 394
 DB 234 CCCCAGATCATCTCTGTCACAAAGTACCCCTCAGACCTCCCGTACATGGGCTCTTAC 175
 QY 395 AlaHisProGlnLeuMetAsnPheValAlaProHisProProAlaLeuProValThrSer 414
 DB 174 GCCACCCACAGAAAGATGACTTGTGGCGCCACCTCCAGCCCTCCCGGACATCT 115
 QY 415 SerSerPhePheAlaIleAlaProAsnProTrpTrpAsnSerProThrGlyGlyIleTrpPro 434
 DB 114 TCCAGTCTTTTGTGCTGCCCAAAACCATGACTGGAATTCACAACTGGGGGTATATACCC 55
 QY 435 AsnThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTrpTrp 451
 DB 54 AACACTAGGCTCCCAACCAAGCCATATGCTTCTCATCTGGGCACTTACTAC 4
 RESULT 15
 ABS14215/C
 ID ABS14215 standard; DNA; 567 BP.
 XX
 AC ABS14215;
 XX
 DT 19-AUG-2002 (first entry)
 XX
 DE Human genome-derived single exon probe ORF from lung SEQ ID No 14206.
 XX
 KW Human; ds; single exon probe; asthma; lung cancer; COPD; ILD;
 KW Chronic obstructive pulmonary disease; interstitial lung disease;
 KW Familial idiopathic pulmonary fibrosis; neurofibromatosis;
 KW Tuberculous scleroderma; Gaucher's disease; Niemann-Pick disease;
 KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
 KW Pulmonary histiocytosis; lymphangioleiomyomatosis; Karsenger syndrome;
 KW Pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;

KW primary ciliary dyskinesia; pulmonary hypertension;
 KW hyaline membrane disease; open reading frame; ORF.
 XX
 XX Homo sapiens.
 OS
 PN WO200106603-A2.
 XX
 PD 15-NOV-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00665.
 XX
 PR 04-FEB-2000; 2000US-180312P.
 PR 26-MAY-2000; 2000US-207456P.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-234687P.
 PR 27-SEP-2000; 2000US-236359P.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 PT WPI; 2002-114183/15.
 DR
 XX Spatially-addressable set of single exon nucleic acid probes, used to
 PT measure gene expression in human lung samples -
 XX
 PA Claim 4; SEQ ID No 14206; 634pp; English.
 XX
 CC The invention relates to a spatially-addressable set of single exon
 CC nucleic acid probes for measuring gene expression in a sample derived
 CC from human lung comprising single exon nucleic acid probes having one of
 CC 12614 nucleic acid sequences mentioned in the specification, or their
 CC complements or the 12387 open reading frames derived from the 12614
 CC probes. Also included are a microarray comprising the novel set of
 CC probes; the novel set of probes which hybridise at high stringency to a
 CC nucleic acid expressed in the human lung; measuring gene expression in a
 CC sample derived from human lung, comprising (a) contacting the array with
 CC a collection of detectably labeled nucleic acids derived from human lung
 CC mRNA, and (b) measuring the label detectably bound to each probe of
 CC the array; identifying exons in a eukaryotic genome, comprising
 CC (a) algorithmically predicting at least one exon from genomic sequences
 CC of the eukaryote; and (b) detecting specific hybridisation of detectably
 CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
 CC having a fragment identical to the predicted exon, the probe is included
 CC in the above mentioned microarray; assigning exons to a single gene,
 CC comprising (a) identifying exons from genomic sequence by the method
 CC above and (b) measuring the expression of each of the exons in several
 CC tissues and/or cell types using hybridisation to a single exon
 CC microarrays having a probe with the exon, where a common pattern of
 CC expression of the exons in the tissues and/or cell types indicates that
 CC the exons should be assigned to a single gene; a peptide comprising one
 CC of 12011 sequences, mentioned in the specification, or encoded by the
 CC probes/open reading frames (ORF). The probes are used for gene
 CC expression analysis, and for identifying exons in a gene, particularly
 CC using human lung derived mRNA and for the study of lung diseases
 CC such as asthma, lung cancer, chronic obstructive pulmonary disease
 CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary
 CC fibrosis, neurofibromatosis, tuberculous scleroderma, Gaucher's disease,
 CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary
 CC haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis,
 CC pulmonary alveolar proteinosis, Karsenger syndrome, fibrocystic
 CC pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension
 CC and hyaline membrane disease. The present sequence is a single exon
 CC probe open reading frame of the invention.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX Sequence 567 BP; 114 A; 124 C; 200 G; 129 T; 0 other;

Alignment Scores:

Pred. No.:	2.91e-71	Length:	567
Score:	948.00	Matches:	169
Percent Similarity:	97.74%	Conservative:	4
Best Local Similarity:	95.48%	Mismatches:	4
Query Match:	38.79%	Indels:	0
DB:	24	Gaps:	0

US-09-902-772-2 (1-451) x ABS14215 (1-567)

```

QY 275 LeuAlaAsnProGlySerGlyInIleGlnLeuTrpGlnPheLeuLeuGlnLeuLeuSer 294
    ||| ::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 534 CTGACCTCATTAGCGAGTGGCCAGATCCAGCTTGGCAGTTCCTCTGAGACTCCTGTG 475

QY 295 AspSerSerAsnSerAsnCysIleThrTrpGlnGlyThrAsnGlyGluPheLysMetThr 314
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 474 GACAGCTCCCACTCCAGCTGCATCACCCTGGGAGGACCAACGGGGAGTTCAAGATGACG 415

QY 315 AspProAspGluValAlaIleArgTrpGlyGluArgLysSerLysProAsnMetAsnTyr 334
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 414 GATCCCGACGAGGTGGCCGGCGCTGGGAGAGCGGAAGACCAACCATGTAACATAC 355

QY 335 AspLysLeuSerArgAlaLeuArgTyrTyrTrpAspLysAsnIleMetThrLysValHis 354
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 354 GATNAGCTCAGCCGCGCCCTCCGTTACTATGACAAAGAACATCATATACCAAGGTCCAT 295

QY 355 GlyLysArgTyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnProHis 374
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 294 GGGAAGCGCTACGGCCTACAACTTGACATTCACCGGATCGCCCAAGCCCTCCAGCCCCAC 235

QY 375 ProProGluSerSerMetTyrLysTyrProSerAspLeuProTyrMetSerSerTyrHis 394
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 234 CCCCCGGAGTCATCTGTACAAAGTACCCCTCAGACCTCCGTTACATGGCTCCTATCAC 175

QY 395 AlaHisProGlnLysMetAsnPheValAlaProHisProProAlaLeuProValThrSer 414
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 174 GCCCAACCCACAGAAAGATGAACCTTGTGTGGGCCCCACCCCTCCAGCCCTCCGTCACATCT 115

QY 415 SerSerPhePheAlaAlaProAsnProTyrTrpAsnSerProThrGlyGlyIleTyrPro 434
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 114 TTCAGTTTTTTTGGTGGCCCAACCATATCGAATTCAACCACTGGGGGTATATATACCCC 55

QY 435 AsnThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr 451
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 54 AACACTAGGCTGCCACGACCATATAGCTTCTCATCTGGGCACCTACTAC 4

```

Search completed: November 9, 2002, 17:14:26
 Job time : 238.257 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 9, 2002, 17:05:02 ; Search time 51.9451 Seconds

(without alignments)
2662.642 Million cell updates/sec

Title: US-09-902-772-2

Perfect score: 2444

Sequence: 1 MASTKEALSVSEDSLEF.....ITYNTRLPAAHPSHLCITY 451

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=framed_p2n.model -DEV=xip
-O=/cgn2_1/USPTO.spool/US0902772/runat.07112002.134034.9055/app.query.fasta.1.1294
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPEXT=0
-LOOPEXT=0 -UNITS=bits START=1 END=1 MATRIX=blosum62 TRANS=human40.cdi
-LIST=43 -DOCALLIG=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US0902772 @CGN_1.1_40 @unat.07112002.134034.9055 -NCPU=6 -ICPU=3
-NO_XLPHY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-MAIN.TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7
-YGAPOP=10 -YGAPEXT=0.5 -DELop=6 -DELEXT=7

Database :

Issued_Patents_NA.*
1: /cgn2_6/ptodata/1/lna/5A.COMB.seq.*
2: /cgn2_6/ptodata/1/lna/5B.COMB.seq.*
3: /cgn2_6/ptodata/1/lna/6A.COMB.seq.*
4: /cgn2_6/ptodata/1/lna/6B.COMB.seq.*
5: /cgn2_6/ptodata/1/lna/PCTUS.COMB.seq.*
6: /cgn2_6/ptodata/1/lna/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	2444	100.0	1447	4	US-08-878-177-1 Sequence 1, Appl1
2	2414.5	98.8	1528	4	US-08-878-177-3 Sequence 3, Appl1
3	1624.5	66.5	2938	2	US-08-343-443B-3 Sequence 3, Appl1
4	511	20.9	1752	4	US-09-360-779-1 Sequence 1, Appl1
5	511	20.9	1752	4	US-09-435-335-1 Sequence 1, Appl1
6	449	18.4	1604	4	US-08-306-691B-43 Sequence 4, Appl1
7	449	18.4	1604	5	PCT-US93-06251-9 Sequence 9, Appl1
8	445.5	18.2	2268	3	US-09-344-579-1 Sequence 1, Appl1
9	342.5	14.0	2667	2	US-08-469-412A-1 Sequence 1, Appl1
10	342.5	14.0	2667	4	US-09-021-715-1 Sequence 1, Appl1
11	322	13.2	2064	3	US-08-875-944B-1 Sequence 1, Appl1
12	322	13.2	2064	4	US-09-116-049-3 Sequence 3, Appl1

13	306	12.5	2410	2	US-08-780-835B-1 Sequence 1, Appl1
14	306	12.5	2410	4	US-09-303-268-1 Sequence 1, Appl1
15	306	12.5	2410	4	US-09-116-049-1 Sequence 1, Appl1
16	301.5	12.3	328	2	US-08-343-443B-5 Sequence 2, Appl1
17	301	12.3	1905	4	US-09-055-113-2 Sequence 5, Appl1
18	291.5	11.9	2266	2	US-09-213-767-1 Sequence 1, Appl1
19	273.5	11.2	5510	3	US-09-009-913-3 Sequence 3, Appl1
20	261.5	10.7	5427	3	US-09-009-913-2 Sequence 2, Appl1
21	261.5	10.7	5667	3	US-09-009-913-4 Sequence 4, Appl1
22	234.5	9.6	2544	2	US-08-469-412A-6 Sequence 6, Appl1
23	234.5	9.6	2544	3	US-09-021-715-6 Sequence 6, Appl1
24	231.5	9.5	848	3	US-09-009-913-38 Sequence 38, App
25	231.5	9.5	2280	3	US-09-009-913-8 Sequence 8, Appl1
26	231.5	9.5	2428	3	US-09-009-913-6 Sequence 6, Appl1
27	231.5	9.5	2498	3	US-09-009-913-10 Sequence 10, Appl1
28	229	9.4	852	4	US-09-020-956-44 Sequence 44, Appl1
29	229	9.4	852	4	US-09-030-607-44 Sequence 44, Appl1
30	229	9.4	852	4	US-09-605-785-44 Sequence 44, Appl1
31	229	9.4	852	4	US-09-439-313-44 Sequence 44, Appl1
32	229	9.4	852	4	US-09-352-616A-44 Sequence 44, Appl1
33	227.5	9.3	852	4	US-09-232-149A-44 Sequence 44, Appl1
34	227.5	9.3	1920	1	US-08-746-789A-1 Sequence 1, Appl1
35	211.5	8.7	2975	1	US-08-368-281-1 Sequence 1, Appl1
36	211.5	8.7	3240	1	US-08-368-281-3 Sequence 3, Appl1
37	181	7.4	65042	4	US-09-784-316-3 Sequence 3, Appl1
38	166	6.8	1364	5	US-08-306-691B-50 Sequence 50, Appl1
39	166	6.8	1364	5	PCT-US93-06251-65 Sequence 65, Appl1
40	118.5	4.8	50937	4	US-09-428-517-1 Sequence 1, Appl1
41	117	4.8	5194	4	US-09-599-652-1 Sequence 1, Appl1
42	117	4.8	5194	2	US-08-642-846-1 Sequence 1, Appl1
43	117	4.8	5194	4	US-09-264-604-1 Sequence 1, Appl1
44	115.5	4.7	2626	1	US-08-156-020-5 Sequence 5, Appl1
45	114.5	4.7	2626	1	US-08-156-020-3 Sequence 3, Appl1

ALIGNMENTS

RESULT 1
US-08-878-177-1
; Sequence 1, Application US/08878177
; Patent No. 6294354
GENERAL INFORMATION:
; APPLICANT: Chugai Selyaku Kabushiki Kaisha, 5001, Iwamoto et
; TITLE OF INVENTION: Cell Calcification Suppressing Proteins and Genes of
; FILE REFERENCE: chugai selyaku kabushiki kaisha 5001
; CURRENT APPLICATION NUMBER: US/08/878,177
; CURRENT FILING DATE: 1997-06-18
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1447
; TYPE: DNA
; ORGANISM: C-11 gene, c-erg gene w/ deletion, chicken DNA
US-08-878-177-1

Alignment Scores:

Pred. No.: 4.05e-247 Length: 1447
Score: 2444.00 Matches: 451
Percent Similarity: 100.00% Conserved: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0

US-09-902-772-2 (1-451) x US-08-878-177-1 (1-1447)

QY 1 MetlaserthrileysglialaleuserValValsergluaBpGlnSerleupheGlu 20
Db 63 ATGCAAGACACTATTAGGAAGCACTTACAGTGTGAGTAAGCAACAGCTCTTGAG 122
QY 21 CyslalarlyrglyserprrohisleualalysrhrGluwetrhralaserSerSerGlu 40
Db 123 TGTGCTTACGGATGCGCCCACTTGCAAAAGACAAATAAGACAGGCTCTTCAGTGA 182

QY 41 TyrGlyGlnThrSerIysMetSerProArgValProGlnGlnAspTrpLeuSerGlnPro 60
 DB 183 TATGGGCAAAATCAAAAGATGAGCCCGCGCTTCCAGAGAGACTGTTATACACAGCC 242
 QY 61 ProAlaArgValThrIleuYsMetGluCysAsnProAsnGlnValAsnGlySerArgAsn 80
 DB 243 CCGGCCAGAGTTTACATTAAGATGAGTGAACCCAAACAGGTAAATGGGTCAAGGAAT 302
 QY 81 SerProAspAspCysSerValAlaIalysGlyIlysMetValSerSerSerAspAsnVal 100
 DB 303 TCACCTGATGACGTCAGCCGTGGCAAAAGAGGAAATGGTTAGCACTTACAGCAATGTT 362
 QY 101 GlyMetAsnTrpGlySerTrpMetGluGluIlysHisIleProProAsnMetThrThr 120
 DB 363 GGGATGAACATGAGAGCTCATGAGAAAGAACATATTCGCCCTCCAAATATGACAAAC 422
 QY 121 AsnGluArgArgValIleValProAlaAspProThrLeuTrpSerThrAspHisValArg 140
 DB 423 AATGAACGAAGAGTTTATGTCAGCAGATCTTACGTTATGAGACACAGCAATGTAAGG 482
 QY 141 GlnTrpLeuGluTrpAlaValIalysGluTrpGlyLeuProAspValAspIleLeuLeuPhe 160
 DB 483 CAGTGGCTGAGTGGGAGAGGAGAGTATGCTTCCACAGCTGGACATCTTGTGTTTC 542
 QY 161 GlnAsnIleAspGlyIysGluLeuCysIysMetThrIysAspAspPheGlnArgLeuThr 180
 DB 543 CAGAACATTCATGGGAAGAGTGTGTAATGACCAAAATGACATCTCCAGACATCAGC 602
 QY 181 ProSerTrpAsnAlaAspIleLeuLeuSerHisIleuHisIleuArgGluValAla 200
 DB 603 CCGAGCTATACCGAGATATCTCTGCTCACCTACCTACCTACCTACGAGAGAAAGAGGCC 662
 QY 201 ThrPheIlePheProAsnThrSerValIlyProGluAlaThrGlnArgIleThrThrArg 220
 DB 663 ACTTTTATTTTCCAAATATACATCAGTTTACCAGAAAGCAAGCAAGAAATACAAACAGG 722
 QY 221 ProAspLeuProTrpGluGlnAlaIleArgSerAlaTrpThrSerHisSerHisProThr 240
 DB 723 CCAAGATTACCTTATGAGCAGAGAGAGATCAAGCGTGGAGTGCACACCATCCCACT 782
 QY 241 GlnSerIysAlaThrGlnProSerSerSerThrValProIysThrGluAspGlnArgPro 260
 DB 783 CACTCAAAAGCTACCAACCATATCTTCAACAGTGCCTCAAAACAGACAGCAGCGCTCT 842
 QY 261 GlnLeuAspProTrpGlnIleLeuGlyProThrSerSerArgLeuAlaAsnProGlySer 280
 DB 843 CAGTTAGATCTTATCAGATTTTGGACCGAGCACAGCAGCGCTTGCAAATTCAGAGGAGT 902
 QY 281 GlyGlnIleGlnLeuTrpGlnPheLeuLeuGluLeuLeuSerAspSerSerAsnSerAsn 300
 DB 903 GGGCAATATACAGTATGAGCGAGTCTTACCTGAGACTTCTGCGAGACAGCTCAACTCCAC 962
 QY 301 CysIleThrTrpGluGluTrpIleAsnGlyIlePheIysMetThrAspProAspGluValAla 320
 DB 963 TGCATACCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1022
 QY 321 ArgArgTrpGluIleArgIysSerIysProAsnMetAsnTrpAspIysLeuSerArgAla 340
 DB 1023 CCGCGTGGGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1082
 QY 341 LeuArgTrpTrpTrpAspIysAsnIleMetThrIysValHisGlyIysArgIlyAlaIlyr 360
 DB 1083 CTTCGCTACTACTATGACAAATAATATTATGACTAAAGTTCATGTAACGCTATGCTTAC 1142
 QY 361 LysPheAspPheHisGlyIleAlaGlnAlaLeuGlnProHisProGluSerSerMet 380
 DB 1143 AATTTGATTTTCCAGGAATGCTCAGGCCCTCAGCCCTCAGCCCTCAGAAATATCATCAG 1202
 QY 381 TyrIysTrpProSerAspLeuProTrpMetSerSerTrpHisAlaHisProGlnIysMet 400
 DB 1203 TACAATACCCATCAGACCTCCCTCATATGAGTTCCTACATGCACACCCCGAGAAAGATG 1262

QY 401 AsnPheValAlaProHisProProAlaLeuProValThrSerSerSerPhePheAlaAla 420
 DB 1263 AACTTGTACTCTCCATCCCTCGCTTTCGCCGTAACCTCATACAGCTTTTGTGCGC 1322
 QY 421 ProAsnProTrpTrpAsnSerProThrGlyIleThrProAsnThrArgLeuProAla 440
 DB 1323 CCAATTCATATGAGATTCACCAACTGGAGGAGATTCACCCCAATACAGGCTGCCAGCT 1382
 QY 441 AlaHisMetProSerHisLeuGlyThrTrpTyr 451
 DB 1383 GCTCATATGCTTCCATCTTGGCACCCTACTAC 1415
 RESULT 2
 US-08-878-177-3
 ; Sequence 3, Application US/08878177
 ; Patent No. 6294354
 ; GENERAL INFORMATION:
 ; APPLICANT: Chugai Selyaku Kabushiki Kaisha, 5001, Iwamoto et
 ; TITLE OF INVENTION: Cell Calcification Suppressing Proteins and Genes of
 ; TITLE OF INVENTION: the proteins
 ; FILE REFERENCE: Chugai Selyaku Kabushiki Kaisha 5001
 ; CURRENT APPLICATION NUMBER: US/08/878,177
 ; CURRENT FILING DATE: 1997-06-18
 ; NUMBER OF SEQ ID NOS: 7
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 3
 ; LENGTH: 1528
 ; TYPE: DNA
 ; ORGANISM: c-erg gene, chicken DNA
 US-08-878-177-3
 Alignment Scores:
 Pred. No.: 5,57e-244 Length: 1528
 Score: 2414.50 Matches: 450
 Percent Similarity: 94.148 Conservative: 0
 Best Local Similarity: 94.148 Mismatches: 1
 Query Match: 98.798 Indels: 27
 DB: Gaps: 1
 US-09-902-772-2 (1-451) x US-08-878-177-3 (1-1528)
 QY 1 MetaIleSerThrIleIysGluAlaLeuSerValIalysGlnAspGlnSerLeuPheGlu 20
 DB 63 ATGGCAAGCACATTAATGAAGAGCATTAATCAGTGTAGTGAACACAGATCTTGTGTGAG 122
 QY 21 CysAlaIlyGlySerProHisIleuAlaIysThrGluMetThrAlaSerSerSerGlu 40
 DB 123 TGTGCTACCGATCGGCCCACTTGCAGAACAGAAATGACAGCTCTCTTCCAGTGAA 182
 QY 41 TyrGlyGlnThrSerIysMetSerProArgValProGlnGlnAspTrpLeuSerGlnPro 60
 DB 183 TATGGGCAAAATCAAAAGATGAGCCCGCGCTTCCAGAGAGACTGTTATACAGGCC 242
 QY 61 ProAlaArgValThrIleuYsMetGluCysAsnProAsnGlnValAsnGlySerArgAsn 80
 DB 243 CCGGCCAGAGTTTACATTAAGATGAGTGAACCCAAACAGGTAAATGGGTCAAGGAAT 302
 QY 81 SerProAspAspCysSerValAlaIalysGlyIlysMetValSerSerSerAspAsnVal 100
 DB 303 TCACCTGATGACGTCAGCCGTGGCAAAAGAGGAAATGGTTAGCACTTACAGCAATGTT 362
 QY 101 GlyMetAsnTrpGlySerTrpMetGluGluIlysHisIleProProAsnMetThrThr 120
 DB 363 GGGATGAACATGAGAGCTCATGAGAAAGAACATATTCGCCCTCCAAATATGACAAAC 422
 QY 121 AsnGluArgArgValIleValProAlaAspProThrLeuTrpSerThrAspHisValArg 140
 DB 423 AATGAACGAAGAGTTTATGTCAGCAGATCTTACGTTATGAGACACAGCAATGTAAGG 482
 QY 141 GlnTrpLeuGluTrpAlaValIalysGluTrpGlyLeuProAspValAspIleLeuLeuPhe 160
 DB 483 CAGTGGCTGAGTGGGAGAGGAGTATGCTTCCACAGCTGGACATCTTGTGTTTC 542

QY 161 GlnAsnIleaspGlyLysGluLeuCySLysMetThrLysaspPheGlnArgLeuThr 180
 Db 543 CAGACATTTGATGGGAAGAGTTGTGTAATAATGACCAAGATGATCTCCAGACTCAGC 602
 QY 181 ProSerThrAsnAlaaspIleLeuSerHisIleuArgGluArg----- 198
 Db 603 CCGAGGTATTAACGAGATATCTCTGTACACCTACATCACTCAGAGACTCTCTT 662
 QY 198 ----- 198
 Db 663 CCACATTTGACTTCAGATGATGTTGATAAGCCTTACAAAACCTCCACGGTTAATGCAT 722
 QY 199 -----GlyAlaThrPheIlePheProAsnThrSerValTyrProGluAla 213
 Db 723 GCTAGAAACACAGAGGAGCCACTTTTATTTTCCAAATACATCAGTTTACCCAGAAAGA 782
 QY 214 ThrGlnArgIleThrArgProAsnLeuProTyrGluGlnAlaArgSerAlaTrp 233
 Db 783 ACGCAAAAGAAATACACAAAGCCAGATTACCTTATGAGCAAGAGAGATCAGCGTGG 842
 QY 234 ThrSerHisSerHisProThrGlnSerLysAlaThrGlnProSerSerThrValPro 253
 Db 843 ACGAGTCACAGCCATCCACATCACTCAAAAGCTACCAACATCATCTTCACAGCTGCC 902
 QY 254 LysThrIleuaspGlnArgProGlnLeuaspProTyrGlnIleLeuGlyProThrSer 273
 Db 903 AAAACACAAAGACCCAGCTCTCTCAGTTAGATCCTTATCAGATTTCTGGACCCAGCAGC 962
 QY 274 ArgLeuAlaaspProGlySerGlyGlnIleGlnLeuThrPheLeuLeuGluLeu 293
 Db 963 CGTTTCGAAATCCAGGAGAGTGGCAGATACAGTATGAGCGTTCCTACCTGGAGCTCTG 1022
 QY 294 SeraspSerAsnSerAsnCysIleThrTrpGluGlyThrAsnGlyGluPheLysMet 313
 Db 1023 TCGGACAGCTCCACATCCACTGCATCACCCTGGAGGAGCAAAATGGGAGTTCAAGATG 1082
 QY 314 ThrAspProaspGluValAlaArgArgTrpGlyLysSerLysProAsnMetAsn 333
 Db 1083 ACAACACCTCATGATGAAGTGGCTGGCGCTGGGAGAGAGAAACCAACCATGTAAC 1142
 QY 334 TyrAspLysLeuSerArgAlaLeuArgTyrTyrAspLysAsnIleMetThrLysVal 353
 Db 1143 TATGACAAACTCAGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1202
 QY 354 HisGlyLysArgTyrAlaTyrLysPheaspPheHisGlyIleAlaGlnAlaLeuGlnPro 373
 Db 1203 CATGGTAAAGCTATGGCTTCAAAATTTGATTCCACGGAATCGCTCAGGCCCTCCAGCCT 1262
 QY 374 HisProGluSerSerMetTyrLysTyrProSeraspLeuProTyrMetSerSerTyr 393
 Db 1263 CACCCCTCCAAATCATCATGATCAATATACCAATACCAACCTCCCTACATGAGTCTTAC 1322
 QY 394 HisAlaHisProGlnLysMetAsnPheValAlaProHisProAlaLeuProValThr 413
 Db 1323 CATGCACACCCCGCAAGAGATGACTTTGAGCTCCCATCCCTGCTGGCCGTAAC 1382
 QY 414 SerSerSerPhePheAlaAlaProAsnProTyrTrpAsnSerProThrGlyLysIleTyr 433
 Db 1383 TCATCCAGCTTTTGTGTCGCCCTAATCCATCATGGAATTCACCAACTGGAGGAGATCTAC 1442
 QY 434 ProAsnThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr 451
 Db 1443 CCCAATACCGAGGCTGCGAGCTGCTCATATGCTTCCATCTTGGCACCTACTAC 1496
 RESULT 3
 US-08-343-443B-3
 ; Sequence 3, Application US/08343443B
 ; Patent No. 5968734
 ; GENERAL INFORMATION:
 ; APPLICANT: Aurias, Alain
 ; APPLICANT: Delattre, Olivier
 ; APPLICANT: Desmaze, Chantal
 ; APPLICANT: Melot, Thomas

; APPLICANT: Peter, Martine
 ; APPLICANT: Ploungastel, Beatrice
 ; APPLICANT: Thomas, Gilles
 ; APPLICANT: Zucman, Jessica
 ; TITLE OF INVENTION: NUCLEIC ACID CORRESPONDING TO A GENE OF
 ; TITLE OF INVENTION: CHROMOSOME 22 INVOLVED IN RECURRENT CHROMOSOMAL
 ; TITLE OF INVENTION: TRANSLATIONS ASSOCIATED WITH THE DEVELOPMENT OF CANCEROUS
 ; TITLE OF INVENTION: TUMORS, AND NUCLEIC ACIDS OF FUSION RESULTING FROM SAID
 ; NUMBER OF SEQUENCES: 129
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Weiser & Associates
 ; STREET: 230 South Fifteenth Street
 ; CITY: Philadelphia
 ; STATE: PA
 ; COUNTRY: USA
 ; ZIP: 19102
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: AEDIT 1.0 DOS text editor
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/343,443B
 ; FILING DATE: 18-NOV-1994
 ; CLASSIFICATION: 514
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/FR93/00494
 ; FILING DATE: 19-MAY-1993
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: FR 92/06123
 ; FILING DATE: 20-MAY-1992
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Weiser, Gerard J.
 ; REGISTRATION NUMBER: 19,763
 ; REFERENCE/DOCKET NUMBER: 989,6121P
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 215-875-8383
 ; TELEFAX: 215-875-8394
 ; INFORMATION FOR SEQ ID NO: 3:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2938 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 143..1498
 ; US-08-343-443B-3
 Alignment Scores:
 Pred. No.: 1,696-160
 Score: 1624.50
 Percent Similarity: 79.43%
 Best Local Similarity: 66.96%
 Query Match: 66.47%
 DB: 2
 Gaps: 7
 US-09-902-772-2 (1-451) x US-08-343-443B-3 (1-2938)
 QY 1 MetAlaSerThrIleLysGluAlaLeuSerValValSerGluAspGlnSerLeuPheGlu 20
 Db 143 ATGACGCGACTATTAAGAGAGCTGTGCGGTGTGAGCGAGCAGCATGCTCTTGTGAC 202
 QY 21 CysAlaTyrGly---SerProHisLeuAlaLysThrGluMetThrAlaSerSerSer 39
 Db 203 TCACGCTACGAGGAGGAGCCCATCTCCCAAGCCGACATGACTGCTCGGAGAGTCT 262
 QY 40 GluTyrGlyGlnThrSerLysMetSerProArgValProGlnGlnAspTrpLeuSerGln 59
 Db 263 GACTACGGGAGCCGCCACAAAGATCAACCCCTCCACACAGCAGAGATGATCAATCAG 322
 QY 60 ProProAlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArg 79

```

Db 323 ---CCAGTGAAGGTCAACGTCAAGCGGAGAT---GACCACATGATGATCCAGG 373
QY AsnSerProAspAspCysSerValAlaLysGlyLysMetValSerSerSerAspAsn 99
Db 374 GAGTCTCCGGTGAAGTCAAGCGGTAGCAAAATGACAGAGGTGTGGGAGGAGGATCC 433
QY 100 ValGlyMetAsnTyrGlySerTyrMetGluGluLysHis---IleProProAsnMet 118
Db 434 AACCCCATCACTCAACCAAGCTATATGACGAGAAAGATGGCCCCCTCTCCCAACATG 493
QY 119 ThrThrAsnGluArgArgValIleValProAlaAspProThrLeuThrSerThrAspHis 138
Db 494 ACCACCAAGAGAGAGAGATCATGTCGCCGACGCCACACACTGTGGACACAGAGCAT 553
QY 139 ValArgGlnTyrLeuGluTyrPalValAlaLysGluTyrGlyLeuProAspValAlaLeu 158
Db 554 GTGAGCATAGGTGGTAGTGGGCGATAAAGAGATAGCTGTATGTAGATGACACATCC 613
QY 159 LeuPheGlnAsnIleAspGlyLysGluLeuCysLysMetThrLysAspPheGlnArg 178
Db 614 TTTTTCACAAACATGATGATGACCAAGCAACTGTAAATATCAACAAAGAGACTTCTCCG 673
QY 179 LeuThrProSerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGluArg 198
Db 674 GCCACCACTCTCAACACAGAGAGTGTGTCTCACACTCAGTTACTCCAGGAA--- 730
QY 199 GlyAlaThrPheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgIleThr 218
Db 731 ---AGTTCACTGCTGCTATATATACACCTCCACACGACCAACATCTCAGCATGAGAT 787
QY 219 ThrArgProAspLeuProTyrGluGlnAlaArgSerAlaThrPheSerHisSerHis 238
Db 788 GTCAAGAAAGACCTTCTTATGACTCAGTCAGAAAGAGAGGAGGCGCATATACAGAAAT 847
QY 239 ProThrGlnSerLysAlaThrGln---ProSerSerSerThrValProLysThrGluAsp 257
Db 848 TCTGGGCTCAACAAAGTCTCTCCCTTGAGAGGCGCAAAACGATCGTAAATACAGAG 907
QY 258 GlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSerSerArgLeuAlaAsn 277
Db 908 CAACGCGCCAGCAGCATCGTATCAGATCCGCGCCGACGACGAGTGGCTGAGCAGCAAC 967
QY 278 ProGlySerGlyGlnIleGlnLeuTyrGlnPheLeuLeuGluLeuLeuSerAspSer 297
Db 968 CCTGGAGGCGGAGATGATCAGCTGTGGCAATCTCTGAGTGTCTCCGACACCGCC 1027
QY 298 AsnSerAsnCysIleThrTyrPgluGlyThrAsnGlyLysPheLysMetThrAspProAsp 317
Db 1028 AACGCCAGCTGATCTACCTGGAGAGGAGACCAACGGGAGATTCAAAATGACGAGCCCGAT 1087
QY 318 GluValAlaArgArgTyrPgluGlyLysSerLysProAsnMetAsnTyrAspLysLeu 337
Db 1088 GAGGTGGCGAGCGCTGGGGGCGAGCAAAACCAACCATGATATACGACAGCTG 1147
QY 338 SerArgAlaLeuArgTyrTyrTyrAspLysAsnIleMetThrLysValHisGlyLysArg 357
Db 1148 ACCCGGGCCCTCGTATATGATGATAAAACATTTATGACCAAAATGACGCGCAAAATA 1207
QY 358 TyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnProHisProGlu 377
Db 1208 TATGCTTCAAAATTGACTTCCACGCGCATGGCCAGGCTCTGACGACCATCCGACCGAG 1267
QY 378 SerSerMetTyrLysTyrProSerAspLeuProTyrMetSerSerTyrHisAlaHisPro 397
Db 1268 TCGTCATGTACAAAGTACCTCTGACATCTCTCATATGCTCTTCACATGCCACGAG 1327
QY 398 GlnLysMetAsnPheValAlaProHisProProAlaLeuProValThrSerSerPhe 417
Db 1328 CAGAAAGTCAACTGTGCTCCCTCCCATCCATCTCATCTCCACTTCCCTCCACACTTC 1387
QY 418 PheAlaAlaProAsnProTyrTyrPheSerProThrGlyLysIleThrProAsn----- 435

```

```

Db 1388 TTGGAGCCGCATCAACATACTGAGACTCCCGCAGGGGGAAATCTAACCCAAACCCCAAC 1447
QY 436 ---ThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr 451
Db 1448 GTCCCGCCGATCTCAACACCCACAGCTGCTTACACTTATAGGAGCTATGAC 1498

RESULT 4
US-09-360-779-1
; Sequence 1, Application US/09360779
; Patent No. 6268216
; GENERAL INFORMATION:
; APPLICANT: Deneris, Evan S.
; APPLICANT: Pyodora, Dmitry V.
; APPLICANT: Hendricks, Timothy J.
; TITLE OF INVENTION: Reagents and Methods for the Screening of Compounds
; FILE REFERENCE: Case-03828
; CURRENT APPLICATION NUMBER: US/09/360,779
; EARLIER FILING DATE: 1999-07-26
; EARLIER APPLICATION NUMBER: 60/094,264
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1752
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (112)..(1131)
US-09-360-779-1

Alignment Scores:
Pred. No.: 7.4e-44 Length: 1752
Score: 511.00 Matches: 123
Percent Similarity: 50.698 Conservative: 23
Best Local Similarity: 42.718 Mismatches: 63
Query Match: 20.918 Indels: 79
Gaps: 10

US-09-902-772-2 (1-451) x US-09-360-779-1 (1-1752)
QY 209 ValTyrProGluAlaThrGlnArgIleThrThrArgProAspLeuProTyrGluGlnAla 228
Db 369 GTGCCCGCCCGCGGCTGCCCGCCCATCTCCACCGCCCA----- 407
QY 229 ArgArgSerAlaThrPheSerHisSerHisProThrGlnSerLysAlaThrGlnProSer 248
Db 408 -----GTCCCGCCGACGATGAGACAGAGCGCACCTCCACGCCCTGCT 452
QY 249 SerSerThrValProLysThrGluAspGlnArg-----Pro 260
Db 453 GATCAACATGTATCTACATCCGATCCGCGAGATGCTTTTAAAGAGAGGAAGAGCC 512
QY 261 GlnLeu-AspProTyrGlnIleLeuGlyProThrSerSerArgLeuAlaAsnProGlySer 280
Db 513 GAGCTGGGGCGG-----CTGAGCCCTCGGTACAGAAA-----GGCAG 551
QY 280 rGlyGlnIleGlnLeuTyrPgluPheLeuLeuGluLeuLeuSerAspSerSerAsnSerAs 300
Db 552 CGGGCAGATCCAGTGTGTGCACTTCTACTGAGCTGCTGGCAGACCGGCGAGCGCG 611
QY 300 nCysIleThrTyrPgluGlyThrAsnGlyLysPheLysMetThrAspProAspGluValAl 320
Db 612 CTGCAATCGGTGGAGAGGGCGGCGAGGCAAGTCAAGCTCACCGACCGCAGAGGTGGC 671
QY 320 AArgArgTyrPgluGlyLysSerLysProAsnMetAsnTyrAspLysLeuSerArgAl 340
Db 672 CGGAGCGTGGGCGAGCGCAAGAGCAAGCCCAATATGAACTACGACAAAGCTAAGTCGAC 731
QY 340 AleuArgTyrTyrTyrAspLysAsnIleMetThrLysValHisGlyLysArgTyrAlaLys 360
Db 732 ACTGCGCTACTACTACGACAAATAATCATATGAGCAAGGTGCGACGGCGACGCTACGCCCTA 791

```


SOFTWARE: WordPerfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/306,691B
FILING DATE: September 15, 1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Monaco, Daniel A.
REGISTRATION NUMBER: 30,480
REFERENCE/DOCKET NUMBER: 8321-8
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-8383
TELEFAX: (215) 568-5549
TELEX: No. 5734039e
INFORMATION FOR SEQ ID NO: 43:
SEQUENCE CHARACTERISTICS:
LENGTH: 1604 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-306-691B-43

Alignment Scores:
Pred. No.: 2,1e-37 Length: 1604
Score: 449.00 Matches: 114
Percent Similarity: 43.88% Conservative: 51
Best Local Similarity: 30.32% Mismatches: 89
Query Match: 18.37% Gaps: 9
DB: 1 92

US-09-902-772-2 (1-451) x US-08-306-691B-43 (1-1604)
QY 119 ThrThrasngluargrValIleValProAlaAspProThrLeuThrSerThrAspHis 138
DB 447 ACTAAAGAACACGACGACTGGGGATCCCAAAAGACCCCGGACGTGACAGAAACCCAT 506
QY 139 ValArgGluTrpLeuGluTrpAlaValLysGluTyrGlyLeuProAspValAspIleLeu 158
DB 507 GTTCGGGACTGGGTGATGTGTGATGTAATTCAGCCCTGAAGAGTAGACTTCAG 566
QY 159 LeuPheGlnAsnIleAspLysLysGluLeuGlySerMetThrLysAspSerPheGlnArg 178
DB 567 AAGTTC---GTATGATGATGACAGCCCTGCGCCCTGGTAAAGACGCTTTCGAG 623
QY 179 LeuThrProSerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgIu--- 197
DB 624 CTGGGCCCGACGACTTCTTGGGACATCTTATGGGAACATCTAGACATCTCGAGAAAGAG 683
QY 198 -----ArgGlyAlaThrPheIlePheProAsnThr----- 207
DB 684 GATGTGAACATATCAATGATGATGAGTCAACCCAGCCTATTCAGAAATCCCGCTATACC 743
QY 207 ----- 207
DB 744 TCGGATTACTTACTAGTACTAGTATGAGCATGCCAGTGTGTCCACCATCGAGTTC 803
QY 208 -----SerValTyrProGluAlaThrGlnArg 216
DB 804 TCAGAGCCCGACTTCATCACAAGTCCATTCAGACCCCTCATCCCATCGAGTCCGGAAGAG 863
QY 217 Ile----- 217
DB 864 CTCTCTCTCCCTCAAGTATGAGATGACATACCCCTGGTCAATTCGCGAGACCCCTTCAG 923
QY 217 ----- 217
DB 924 ACAGACACCTTGCAAGATGACTACTTGTCTATCAACAAGAAAGTGTCAACCCCGAGAAC 983
QY 218 -----ThrThrArgProAspLeuProTyrGluGluAlaIleArgSer--- 231
DB 984 ATGTGATGGGAGGAGACCAAGTGTGTAAACTCGGGGGCCAGGACTCTTTTGAAGACATA 1043

QY 232 ---AlaTrpThrSerHisSerHisProThrGlnSerLysAlaThrGlnProSerSer 250
DB 1044 GAGAGCTACGATAGTGTGATGTCGCTCACCCAGCTCTGGAGCAGGACATCATCTTCAC 1103
QY 251 ThrValProLys-----ThrGlnAsp----- 257
DB 1104 AGCCTGACGGTGTCCCTCATATGACAGCTTCGACTCAGAGAGCATATCCGCTGCCCTG 1163
QY 258 -----GlnArgProGlnLeuAsp 263
DB 1164 CCCAACCAAGCCCAAGGACACTTCAGAGACTATGTGGGAGCCGTGCTGACTCAAT 1223
QY 264 ProTyrGlnIleLeuGlyProThrSerSerArgLeuAlaAsnProGlySerGlyGlnIle 283
DB 1224 AAGCAACAGCCCTCATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1283
QY 284 GlnLeuTrpGlnPheLeuLeuGluLeuLeuLeuSerAspSerSerAsnSerAsnGlyIle 303
DB 1284 CAGCTATGGCAGTTCCTTCTGGAATTCACACATCAATCCATGATAAATCCGTGACGCTTTATCAGC 1343
QY 304 TrpGluGlyThrAsnGlyLuphLeuPheLysMetThrAspProAspGluValAlaArgArgTrp 323
DB 1344 TGGACAGAGATGCTGCGGAATTCATCAACTTCTTGACCCAGATGAGGTGGCCAGAGATGG 1403
QY 324 GlyLysArgLysSerLysProAsnMetAsnTyrAspLysLeuSerArgAlaLeuArgTyr 343
DB 1404 GGAAGAGAGAAAAACAACCTAGATGATGATTAAGAAACTGAGCGGTGGCTGACGCTAC 1463
QY 344 TyrTyrAspLysAsnIleMetThrLysValHisGlyLysArgTyrAlaTyrLysPhe--- 362
DB 1464 TATTTACGACAAAAACATCATCCACAGACGCGGGAACCGCTACCTGATCGCTTTGTGG 1523
QY 363 ---AspPheHisGlyIleAlaGlnAlaLeuGlnProHisProProGlu 377
DB 1524 TGTGACCTG-----CAGACCTGCTGGGTATACACCCCTGAG 1559

RESULT 7
PCT-US93-06251-9
Sequence 9, Application PC/TUS9306251
GENERAL INFORMATION:
APPLICANT: Wickstrom, Eric and Rife, Jason P.
TITLE OF INVENTION: Trivalent Synthesis of Oligonucleotides Containing
NUMBER OF SEQUENCES: 93
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCUDLY, SCOTT, MURPHY & PRESSER
SMREER: 400 Garden City Plaza
CITY: Garden City
STATE: NY
COUNTRY: USA
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/06251
FILING DATE: 19930630
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Digilio, Frank S.
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 8586
TELECOMMUNICATION INFORMATION:
TELEPHONE: 516-742-4366
TELEFAX: 516-742-4343
TELEX: 230 901 SANS UR
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1604 base pairs
TYPE: nucleic acid

;	STRANDEDNESS:	double
;	TOPOLOGY:	Linear
;	MOLECULE TYPE:	DNA (genomic)
;	PCT-US93-06251-9	
Alignment Scores:		
Pred. No.:	2,1e-37	Length: 1604
Score:	449.00	Matches: 114
Percent Similarity:	43.88%	Conservative: 51
Best Local Similarity:	30.32%	Mismatches: 89
Query Match:	18.37%	Indels: 122
DB:	5	Gaps: 9
US-09-902-772-2 (1-451) x PCT-US93-06251-9 (1-1604)		
QY	119	ThrThraAngIaTgArGValIleValProAlaAspProThrLeuTrpSerThraSphHis 138
DB	447	ACTAAAGACAGACAGACAGCTGGGGATCCCAAAAGACCCCGGACAGACAGAAACCCAT 506
QY	139	ValArgInTrpLeuGluTrpAlaValLySgluTyGlyLeuProAspValSphIleu 158
DB	507	GTTCCGGACCTGGGGATGATGGCCCTGGCAAGAAATTCACGCTGAAGGTGTACCTTCACG 566
QY	159	LeuPheGlnAsnIleAspGlyLySgluLeuCySlyMetThrLySAspSphPheGlnArg 178
DB	567	AAGTTC---TGTATGATGTGAGACAGCCCTTGGCGCCCTGGGTAAACACTGTTTCTCCAG 623
QY	179	LeuThrProSerTyTrpAsnAlaAspIleLeuLeuSerHisLeuHisTyTrpLeuArgIu--- 197
DB	624	CTGGCCCCAGACTTGTGTGGGACATCTTATGGGAACATCTAGACATCTCGCAGAAAGAG 683
QY	198	-----ArgGlyAlaThrPheIlePheProAsnThr----- 207
DB	684	GATGTGAACCATATCATCAAGTTAAATGGAGTCCACCCAGCCTATCCAGAAATCCCGCTATACC 743
QY	207	----- 207
DB	744	TCGGATTACTTCATTAGCTATGTTATGAGCATGCCCACTGTGTTCCACCATCGAGTTTC 803
QY	208	-----SerValyTrpProGluIatThrGlnArg 216
DB	804	TCAGAGGCCAGCTTCATCACAGAGTCCATACAGACCGTCCATCCATCCATCAGCTCGGAAGAG 863
QY	217	Ile----- 217
DB	864	CTCCTCTCCCTCAAGTATGAGAAATGACTACCCCTCGCTGCTATTCCTCGAGACCGCTCCAG 923
QY	217	----- 217
DB	924	ACAGACACCTTGCAGATGACTACTTTGCTATCAACAGAAAGTCTCACCCACAGACAAAC 983
QY	218	-----ThrThrArgProAspLeuProTyGlnGlnIatIlaArgIser--- 231
DB	984	ATGTGCATGGGGAGGACCACTGCTGGTAAATCTCGGGGCCAGGACTCTTTTGAAAGCATATA 1044
QY	232	---AlatTrpThrSerHisIserHisProThrGlnIserLyAlaThrGlnProSerSerSer 250
DB	1044	GAGAGCTACGATAGTTGTGTGATCCGCTACCCACCACTCCTGGAGGAGCCAGATCATCTTCAAC 1104
QY	251	ThrValProIlys-----ThiGluAsp----- 257
DB	1104	AGCCTGAGGGGTGTTCCTCTCTATGACAGCTTCGACAGAGACATATCCGGCTGGCCCTG 1164
QY	258	-----GlnArgProGlnLeuAsp 263
DB	1164	CCCAACACAAAGCCCAAGGCACTTCACAGACTATGTGCGGGACCGTCTGACCTCAAT 1224
QY	264	ProTyGlnIleLeuGlyProThrSerSerArgLeuAlaAspProGlySerGlyGlnIle 283
DB	1224	AAGGACAAGCGTCGATCTCTGCTGGTGGCCCTTACGTGGCTACACAGACAGCTGGACCAATTC 1284
QY	284	GlnLeuTrpGlnPheLeuLeuGluLeuLeuLeuSerAspSerSerAsnSerAsnGlyIleThr 303

[illegible]


```

Db 427 TCACCTACAAATTCATTTCAACAACCTGCTGCTCATTTACCCATTCATTCATGTCG 486
Qy 366 -----GlyllealaglnaleuInProHisProGlnSerMet 381
Db 487 GGTGGCTGGGGGTGACAGTCCACAGAGTCCCGCCAGTCCGCGGGGTAGCCACT 546
Qy 381 YLYSTYrProSerAspLeuProTYrMetSerSerTYrHisAlaHisProGlnLysMet 401
Db 547 TCCGCTTCCTCCCTCAACGCC-----TCCG 573
Qy 401 snphevalaPro-----HisProAlaLeuProValThrSerSers 416
Db 574 AGGTCGTCTCCCGCCACGAGACCCCGCTCACCACACGCTCTTCATTCATTCATCTT 633
Qy 416 erphepheaAla 420
Db 634 CCTCTCTCGGCT 647

RESULT 10
US-09-021-715-1
; Sequence 1, Application US/09021715
; Patent No. 6194547
; GENERAL INFORMATION:
; APPLICANT: MayoThalasit's, George J.
; Blair, Donald G.
; Fisher, Robert J.
; Beal Jr., Gregory J.
; Athanasios, Meropi A.
; Sgouras, Dionysios N.
; TITLE OF INVENTION: The ERF Genetic Locus and Its Products
; CORRESPONDENCE ADDRESSES: 16
; NUMBER OF SEQUENCES: 16
; ADDRESS: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/021,715
; FILING DATE: 10-Feb-1998
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Garrett-Wackowski, Eugenia
; REGISTRATION NUMBER: 37,330
; REFERENCE/DOCKET NUMBER: 015280-229000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2667 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 123..1769
; OTHER INFORMATION: /note= "human ERF (ETS2 Repressor
; Factor) cDNA"
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
; US-09-021-715-1

Alignment Scores:
Pred. No.: 7.49e-26 Length: 2667
Score: 342.50 Matches: 84
Percent Similarity: 53.40% Conservative: 26

```

```

Best Local Similarity: 40.78% Mismatches: 59
Query Match: 14.01% Indels: 38
DB: 4 Gaps: 5

US-09-902-772-2 (1-451) x US-09-021-715-1 (1-2667)
Qy 239 ProthrgInSerLysAlaThrGlnProSerSerSerThrValProLysThrGlnAsp--G 258
Db 74 CCGGAGACCGCGGCCCGGAAATCGGGGCGCTTCGCCCGCGGCCCGCCACGATGAAAGACCC 133
Qy 258 InArgProGlnLeuAspProTYrGlnIleLeuGlyProHisSerSerArgLeuAlaAsp 278
Db 134 GCGGACACAGAGGTTGGCTTCCCGGATT--GGGCTTACA--AGCCAGAG-TGCTCC 186
Qy 278 roGlySerGlyGlnIleGlnLeuThrGlnPheLeuLeuGlnLeuLeuSerAspSerSera 298
Db 187 CTGGCTCAAGCAGAGATCCAGCTGTGACCTTTATCTTGAGTGTGCGGAGAGAGAGT 246
Qy 298 snSerAsnCysIleThrTrpGluGlyThrAsnGlyGluPheLysMetThrAspProAspG 318
Db 247 ACCAGGCGCTATTCCTGCGCAGGGGAGTACGGGGAATTCGTCATCAAGACCTGATG 306
Qy 318 luValAlaArgArgTrpGlyGluArgLysSerLysProAsnMetAsnTYrAspLysLeuS 338
Db 307 AGGTGCCCCGCTGTGGGGCGCTCCGCAAGTCGACAGCCAGATTAATTCAGCAAGCTGA 366
Qy 338 erArgAlaLeuArgTrpTYrTYrAspLysAsnIleMetThrLysValHisGlyLysArgT 358
Db 367 GCGGGGCGCTCGCTATTCATCTTACAAAGCGCATTCGACACAGACCAAGGGGAAACGCT 426
Qy 358 YrAlaTYrLysPheAspPheHis----- 365
Db 427 TCACCTACAAATTCATTTCAACAACCTGCTGCTCATTCATTCATTCATTCATTCATG 486
Qy 366 -----GlyllealaglnaleuInProHisProGlnSerMet 381
Db 487 GGTGGCTGGGGGTGACAGTCCACAGAGTCCCGCCAGTCCGCGGGGTAGCCACT 546
Qy 381 YLYSTYrProSerAspLeuProTYrMetSerSerTYrHisAlaHisProGlnLysMet 401
Db 547 TCCGCTTCCTCCCTCAACGCC-----TCCG 573
Qy 401 snphevalaPro-----HisProAlaLeuProValThrSerSers 416
Db 574 AGGTCGTCTCCCGCCACGAGACCCCGCTCACCACACGCTCTTCATTCATTCATCTT 633
Qy 416 erphepheaAla 420
Db 634 CCTCTCTCGGCT 647

RESULT 11
US-08-875-944B-1
; Sequence 1, Application US/08875944B
; Patent No. 6096542
; GENERAL INFORMATION:
; APPLICANT: FUJINAGA, Kei
; APPLICANT: YOSHIDA, Koichi
; APPLICANT: HIGASHINO, Fumihiko
; TITLE OF INVENTION: CANCER CONTROL
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.
; STREET: 624 Ninth Street N.W., Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

```



```

; LENGTH: 2410 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-780-835B-1

```

Pred. No.:	4.35e-22	241
Score:	306.00	109
Percent Similarity:	36.67%	34
Best Local Similarity:	27.95%	107
Query Match:	12.52%	141
DB:	2	14
	Gaps:	14

US-09-902-772-2 (1-451) x US-08-780-835B-1 (1-2410)

QY	45	SetrYsMetSerProAsgValProGlnIAspRrPreLeuSerGlnProProAlaArgVal	64
Db	634	TCAAGAAAGAAACCCCAAGAGTCCC-----GCACAGACCCCGGCTGCT	67B
QY	65	ThrIleuYsMetGluCysAsnProAsnGlnValAsnGlySerAlaArgAsnSerProAsp--	83
Db	679	GCACAGAGAAAGCCACACTCCCTCCACACATGAGAGAGAGAGCTTACTCTTACCTGACAA	738B
QY	84	-----AspCysSerValAlaIAs	89
Db	739	TGCGCATCATAGTCCCGGCTCCGGTCCCGCTGGACAGTGGCCCTCCAGCGCTTTTCA	798B
QY	90	GlyGlyLysMetValSerSerSerAspAsnValGlyMetAsnTyrGlySerTyrMetGlu	109B
Db	799	GGCGAGAAACAGCAGCAGAGCTCC-----TGACAG	828B
QY	110	GlnLysHisIleProProProAsnMetThrThiAsnGlnArgArgValIleValProIa	123B
Db	829	CTCCAGACCTCTCCACATCCACCTCCAGCTGGACAGGGTACTGTGTGACACAGCTCCCTCT	888B
QY	130	AspProThrLeuTyrSerThrAspHisValArgGlnTyrPreuGlnTyrAlaValGlu	149B
Db	889	TCCAGCAGCCCGGTGACA-----	906B
QY	150	TyrGlyLeuProAspValAspIleLeuLeuPheGlnAsnIleAspGlyLysGlnLeuCys	165B
Db	907	-----TGT	909B
QY	170	LysMetThrLysAspArgPheGlnArgLeuThrProSerTyrTrpAsnIleAspIleLeu	189B
Db	910	-----GCCATCTCTTCA	921B
QY	190	SerHisLeuHisTyrLeuArgGluArgGlyAlaThrPheIlePhePro-----AsnThr	207B
Db	922	-----CATCTCCAGGAGAGGGGGCGGAGAACTCTCCAGCCGCCCTATCATACACC	972B
QY	208	Ser-----ValTyrProGlnAlaThrGlnArg-----	216B
Db	973	AACGTGTGGAGCCCTGCCACACCTTACCCAGCAGAACTTCAAGCAGAGTACCATTGACC	103B
QY	217	---IleThrThiArgProAspLeuProTyrGlnGlnIleAlaArgAspSerIa-----	232B
Db	1033	CCCTGTACAGAAACGGCTGGCCAGCCGCT---CAAGCCAGGGTGGGGTCAAGTGGGACAA	108B
QY	233	-----TrrPheSerHisSerHisIleProThrGlnSerLysAlaThrGln	246B
Db	1090	GGTACCCAGAGGGCGGGGTGTGATCAAAACAGAGAGCCACAGACTTGCTGACTAGACTCAG	1144B
QY	247	ProSerSerSerThrValProLysThrGlnAspGlnArgProGlnIleuAspProTyrGln	266B
Db	1150	ATGTCTCTGGATGTGCATCAATGTACTCCACACAGAGGGCTTCTGTGAGCCCTTCAG	120B
QY	267	Ile-----	267B
Db	1210	GTATGTAGAGTATGGGTATATGGCATGAATAATCCCTTGACACCATTCACAGATGATCT	126B
QY	268	-----LeuGlyProIleHisSerArg-----LeuAlaAsn	277B

Db	1270	GCATGTCCCTAAAAAATTGGAGGAGACATCAGCAGGAAAGCATTTGGAGCTTCCGG	1329
QY	278	ProGly-----SerGlyInIleGlnLeuTrpGlnPheLeuLeuLeu	292
Db	1330	AG-GGGCACCCCTACCAGCGCCGGCGTGGCTTACAACTGGCAGCTTCTGGTGGCCCTG	1388
QY	293	LeuSerAspSerSerAsnSerAsnGysIleHrTrpGluGlyPheAsnGlyGluPheLys	312
Db	1389	CTGATCAACCCAAACAAATGGTCTATTTCATTCCTGGACAGCGCGGGGAATGGAGTTTAA	1448
QY	313	MetThrAspProAspGluValAlaIleArgArgTrpGlyGluArgLysSerLysProAsnMet	332
Db	1449	CTATTGTAACCTGAAGAGGTGGCCAGCGCTGGGGATGCACAAAGAACCGCCACCCATG	1508
QY	333	AsnTrpAspLysLeuSerAlaGluLeuArgTrpTrpTrpAspAsnIleMetThrLys	352
Db	1509	AATTATGACAACTGAGCGCGCTGCGCATCTATTATGACAAAGCATCATGCAGAAAG	1568
QY	353	ValHisGlyLysArgTrpAlaArgLysPhe	362
Db	1569	GTGGCTGGCGAAGCGCTACAGTGTCACAAAGTTT	1598

```

RESULT 14
US-09-303-268-1
: Sequence 1, Application US/09303268
: Patent No. 6172212
:
GENERAL INFORMATION:
: APPLICANT: Hung, Mien-Chle
:           Xing, Xiangming
: TITLE OF INVENTION: PEAS Is a Tumor Suppressor
: NUMBER OF SEQUENCES: 9
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: ARNOLD, WHITE AND DURKEE
: STREET: P.O. Box 4433
: CITY: Houston
: STATE: Texas
: COUNTRY: USA
: ZIP: 77210-4433
:
COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
:
CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/303,268
: FILING DATE: 30-Apr-1999
: CLASSIFICATION: <Unknown>
:
PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/780,835
: FILING DATE: 10-JAN-1997
:
ATTORNEY/AGENT INFORMATION:
: NAME: Wilson, Mark B.
: REGISTRATION NUMBER: 37,259
: REFERENCE/DOCKET NUMBER: UTSC500
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (512) 418-3000
: TELEFAX: (512) 474-7577
:
INFORMATION FOR SEQ ID NO: 1:
:
SEQUENCE CHARACTERISTICS:
: LENGTH: 2410 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
:
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
:
US-09-303-268-1
:
Alignment Scores:
: Pred. No.: 4.35e-22 Length: 2410
: Score: 306.00 Matches: 109
: Percent Similarity: 36.67% Conservative: 34
: Best Local Similarity: 27.95% Mismatches: 107
: Query Match: 12.52% Indels: 141

```



```
Db 922 -----CATCTCTCAGAGGAGGGGCGGAGACCTCTCCAGCCCTATCAACACC 972
QY 208 Ser-----ValTyrProGluAlaThrGlnArg----- 216
    ::::
Db 973 AACTGTGGAGCCCTGGCCACCTACCCAGAGAACTCAAGCAGAGATACCATGACC 1032
QY 217 ---IleThrThrArgProAspLeuProTyrGluGlnAlaArgArgSerAla----- 232
    ::::
Db 1033 CCCTGTACGAAAGCTGCGCCAGCTT--CAAGCCAGGCTGGGGTCACTGGGCACA 1089
QY 233 -----TrrThrSerHisSerHisProThrGlnSerLysAlaThrGln 246
    ::::
Db 1090 GGTACCCAGAGGGGGGGGGGTGTATCAAAAGAGAGCGCACAGACTTCGCTACGACTCAG 1149
QY 247 ProSerSerSerThrValProLysThrGlnAspGlnArgProGlnLeuAspProTyrGln 266
    ::::
Db 1150 ATGTCCCTGGATGTGATCATGTACCTCCACCCAGAGGGGCTTCTCTGACCCCTCCAG 1209
QY 267 Ile----- 267
    ::::
Db 1210 GTGATGGAGTGANGGTTATGGCTATGAAATAATCCCTTCGACCATTCACGATGATGCTT 1269
QY 268 -----LeuGlyProThrSerSerArg-----LeuAlaAsn 277
    ::::
Db 1270 GCATGTCCCTAAAAAATTGAAAGAGACATCAAGCAGAGAGGATTTGGAGCTTTCCGGG 1329
QY 278 ProGly-----SerGlyGlnIleGlnLeuThrGlnPheLeuLeuGlnLeu 292
    ::::
Db 1330 AG-GGGCCACCTACAGCGCCGGGGTGCCTTACACTGTGCGAGTTTGTGGCCCTG 1388
QY 293 LeuSerAspSerSerSerSerAsnGlyIleThrTyrGlnGlyThrAsnGlyGluPheLys 312
    ::::
Db 1389 CTGGATGACCCCAACAAATGCTCATTTTCATTGCTTGGACAGCGCGGGAATGAGATTAA 1448
QY 313 MetThrAspProAspGluValAlaArgArgTyrGlnArgLysSerLysProAsnMet 332
    ::::
Db 1449 CTAATTGAACCTGAAGAGGTGTCAGGCTCTGGGGTATCCAGAAAGACGGCCAGCCATG 1508
QY 333 AsnTyrAspLysLeuSerArgAlaLeuArgTyrTyrTyrAspLysAsnIleMetThrLys 352
    ::::
Db 1509 AATTATGACAGCTGAGCGCTGCTGCTACTATTATGAGAAAGCATCATGCAAGAG 1568
QY 353 ValHisGlyLysArgTyrAlaTyrLysPhe 362
    ::::
Db 1569 GTGGCTGGCAAGCTACGTACAAAGTTT 1598
    ::::
```

Search completed: November 9, 2002, 19:48:56
Job time : 69.9451 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 9, 2002, 18:55:38 ; Search time 65.0527 Seconds

(without alignments)
2459:669 Million cell updates/sec

Title: US-09-902-772-2

Perfect score: 2444
Sequence: 1 MASTIKKALSVSEDSQSLFE.....IYPNTRLPAHMPHSLGTY 451

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 320260 segs, 177392727 residues

Total number of hits satisfying chosen parameters: 640520

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODE=frame+ p2n.model -DEV=xlp
-O/cgnt2.1/USPTO.spool/US09902772/runat.07112002.134035.9131/app.query.fasta.1.1294
-DB=Published.Applications.NA -GENE=fastap -SUFFIX=trmp MINMATCH=0.1
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsun62
-TRANS=human40.csl -LIST=45 -DOCALIGN=200 -THR_SCORE=pcr -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09902772.ecgn.1.146.ecgnat.07112002.134035.9131
-NCPV=6 -ICPV=3 -NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published.Applications.NA:*
1: /cgnt2.6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgnt2.6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgnt2.6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgnt2.6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
5: /cgnt2.6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
6: /cgnt2.6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
7: /cgnt2.6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
8: /cgnt2.6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
9: /cgnt2.6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgnt2.6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgnt2.6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgnt2.6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgnt2.6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgnt2.6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Length	ID	Description
1	948	38.8	567 10	US-09-864-761-18410
2	863	35.3	473 10	US-09-864-761-20472
3	712	29.1	549 10	US-09-923-779-54
4	650	26.6	420 10	US-09-864-761-3705

Result No.	Query Match	Length	ID	Description
5	650	26.6	454 10	US-09-864-761-1653
6	511	20.9	1752 10	US-09-850-799-1
7	444.5	18.2	2188 10	US-09-920-300A-1716
8	444.5	18.2	2188 12	US-10-033-528-1716
9	444.5	18.2	2268 10	US-09-920-300A-1693
10	444.5	18.2	2268 12	US-10-033-528-1693
11	431.5	17.7	553 10	US-09-925-300-420
12	374	15.3	1884 10	US-09-864-761-7427
13	371	15.2	225 10	US-09-864-761-21761
14	371	15.2	348 10	US-09-864-761-29364
15	371	15.2	477 10	US-09-864-761-5035
16	371	15.2	519 10	US-09-864-761-12798
17	339	13.9	3178 9	US-10-108-605-124
18	339	13.9	3178 9	US-10-108-605-128
19	322	13.2	2333 10	US-09-920-300A-1788
20	322	13.2	2333 10	US-09-880-107-3316
21	322	13.2	2333 12	US-10-033-528-1788
22	301	12.3	1894 10	US-09-126-945B-1
23	301	12.3	1905 10	US-09-866-356-2
24	300.5	12.3	2952 10	US-09-925-297-257
25	280	11.5	1894 10	US-09-841-963A-1
26	279	11.4	165 10	US-09-864-761-23294
27	260.5	10.7	1429 10	US-09-764-864-320
28	255	10.4	1426 10	US-09-925-297-309
29	247	10.1	1704 10	US-09-841-963A-3
30	235.5	9.6	4190 10	US-09-962-832-111
31	231.5	9.5	1915 10	US-09-964-824A-101
32	231.5	9.5	1915 10	US-09-964-824A-563
33	231.5	9.5	1915 10	US-09-880-107-3420
34	231.5	9.5	1915 10	US-09-967-768A-192
35	231.5	9.5	1917 10	US-09-922-217-1105
36	231.5	9.5	1966 10	US-09-925-301-207
37	229	9.4	852 10	US-09-759-143-44
38	229	9.4	852 10	US-09-764-669-44
39	229	9.4	852 10	US-09-030-606-44
40	229	9.4	852 10	US-09-822-827-44
41	229	9.4	852 10	US-09-822-827-44
42	205.5	8.4	852 10	US-09-115-453-44
43	204.5	8.4	488 10	US-09-777-564-468
44	204.5	8.4	507 10	US-09-864-761-27055
45	194	7.9	2642 10	US-09-777-564-489
				US-09-920-300A-1680

ALIGNMENTS

RESULT 1
US-09-864-761-18410/c
Sequence 18410, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OR INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aeomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263,6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30

```

; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 18410
; LENGTH: 567
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AP000163.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.92
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN B1474, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.95
; OTHER INFORMATION: SWISSPROT HIT: P11308, EVALUE 9.00e-99
; OTHER INFORMATION: EST_HUMAN HIT: R87572.1, EVALUE 0.00e+00
; OTHER INFORMATION: NT HIT: M17254.1, EVALUE 0.00e+00
US-09-864-761-18410

Alignment Scores:
Pred. No.: 2,12e-86 Length: 567
Score: 948.00 Matches: 169
Percent Similarity: 97.74% Conservative: 4
Best Local Similarity: 95.48% Mismatches: 4
Query Match: 38.79% Indels: 0
DB: 10 Gaps: 0

US-09-902-772-2 (1-451) x US-09-864-761-18410 (1-567)
QY 275 LeuAlaAsnProGlySerLysGlnIleGlnLeuTrpGlnPheLeuLeuLeuSer 294
DB 534 CtgAGCTCTATGAGCGATGCGGAGATCCAGCTTGGCACTTCCTGAGCTCCGTG 475
QY 295 AspSerSerAsnSerAsnGlySileThrTrpGluGlyThrAsnGlyGluPheLysMetThr 314
DB 474 GACAGCTCCAACTCCAGCTGCATCACTGGGAAGGACCAAGGGAGTTCAAGATGAGC 415
QY 315 AspProAspGluValAlaAlaArgArgTrpGluArgLysSerLysProAsnMetAsnTyr 334
DB 414 GATCCCGAGAGAGGTGGCGCGGCGTGGGAGAGCGGAAGACCAACCAATGAACTAC 355
QY 335 AspLysLysSerArgAlaLeuArgTyrTyrTyrAspLysAsnIleMetThrLysValHis 354
DB 354 GATTAAGCTCAGCGCGCCCTCCGTTACTACTATGACAAACATCATCATGACCAAGTCCAT 295
QY 355 GlyLysArgTyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnProHis 374
```

```

DB 294 GGGAAAGCCGCTACCCCTACAGATTGCACCTTCCACGGGATGCGCCAGCCCTCCAGCCAC 235
QY 375 ProProGluSerSerMetLysTyrProSerAspLeuProTyrMetSerTyrHis 394
DB 234 CCCCCGAGATCATCTCTGTACAGTACCCTCGACCTCCGATGAGGCTCTATAC 175
QY 395 AlaHisProGlnLysMetAsnPheValAlaProHisProProAlaLeuProValThrSer 414
DB 174 GCCCCACCCAGAAAGATGAACTTGTGGGCCCCCACCCTCCAGCCCTCCGACATCT 115
QY 415 SerSerPhePheAlaAlaProAsnProTyrTrpAsnSerProThrGlyLysIleTyrPro 434
DB 114 TCCAGTTTTCCTGCTCCCAAAACCATACTGGAATTCACCAACTGGGGTATATACCC 55
QY 435 AsnThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr 451
DB 54 AACACTAGGCTCCCGACAGCATATGCTTCATCTGAGGCACTTACTAC 4

RESULT 2
US-09-864-761-20472/c
; Sequence 20472, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Aegonica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 20472
; LENGTH: 473
; TYPE: DNA
```



```

: PRIOR APPLICATION NUMBER: PCT/US01/00666
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00667
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00664
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00669
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00665
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00668
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00663
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00662
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00661
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00670
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: US 60/234,687
: PRIOR FILING DATE: 2000-09-21
: PRIOR APPLICATION NUMBER: US 09/608,408
: PRIOR FILING DATE: 2000-06-30
: PRIOR APPLICATION NUMBER: US 09/774,203
: PRIOR FILING DATE: 2001-01-29
: NUMBER OF SEQ ID NOS: 49117
: SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
: SEQ ID NO 3705
: LENGTH: 420
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: OTHER INFORMATION: MAP TO AP000021.2
: OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.2
: OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1
: OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1
: OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1
: OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.97
: OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.5
: OTHER INFORMATION: EXPRESSED IN HL100, SIGNAL = 1.4
: OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1
US-09-864-761-3705

Alignment Scores:
Pred. No.: 9.59e-57 Length: 420
Score: 650.00 Matches: 114
Percent Similarity: 98.31% Conservative: 2
Best Local Similarity: 96.61% Mismatches: 2
Query Match: 26.60% Indels: 0
DB: 10 Gaps: 0

US-09-902-772-2 (1-451) x US-09-864-761-3705 (1-420)
QY 334 TyrAspLysIleuSerArgAlaIleuArgTyrTyrTyrAspLysAsnIleMetThrLysVal 353
Db 418 TACGATTAAGCTGACGGCGCCCTCCGCTACTACTATGACAAAGAAATCATGACCAAGGTC 359
QY 354 HisGlyLysArgTyrAlaTyrLysPheAspPheHisGlyIleAlaGlnPro 373
Db 358 CATGGAGAGCGCTACAGTTCAGCTTCACGGGAGATCGCCAGGCCCTCCAGGCC 299
QY 374 HisProPogIuSerSerMetTyrLysTyrProSerAspLeuProTyrMetSerTyr 393
Db 298 CACCCCGGAGTCACTCTGTACAAAGTACCCCTCAGACCTCCGCTACATGGGCTCTAT 239
QY 394 HisAlaHisProGlnLysMetAsnPhaValAlaProHisProProAlaIleuProValThr 413
Db 238 CACGCCCAACCAACAAGATGACTTGTGGGGCCCAACCTCCAGCCCTCCGCTGACA 179
QY 414 SerSerSerPhePheAlaAlaProAsnProTyrTyrPasnSerProThrGlyIleTyr 433
Db 178 TCTTCGACTTTTTCGTCGCCCAACCAACCATACCTGGAATTCACCAACAGCGGGGTATATAC 119
```

```

QY 434 ProAsnThrArgLeuProAlaHisMetProSerHisLeuGlyThrTyrTyr 451
Db 118 CCCAACACTAGGCTCCCAAGCAGCATATGCTTCTCATCTGAGGCACTTACTAC 65

RESULT 5
US-09-864-761-1653/c
: Sequence 1653, Application US/09864761
: Patent No. US20020048763A1
: GENERAL INFORMATION:
: APPLICANT: Penn, Sharon G.
: APPLICANT: Rank, David R.
: APPLICANT: Hanzel, David K.
: APPLICANT: Chen, Wensheng
: TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
: FILE REFERENCE: Aecmica-x-1
: CURRENT APPLICATION NUMBER: US/09/864,761
: PRIOR APPLICATION NUMBER: US 60/180,312
: PRIOR FILING DATE: 2000-02-04
: PRIOR APPLICATION NUMBER: US 60/207,456
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: US 09/632,366
: PRIOR FILING DATE: 2000-08-03
: PRIOR APPLICATION NUMBER: GB 24263,6
: PRIOR FILING DATE: 2000-10-04
: PRIOR APPLICATION NUMBER: US 60/236,359
: PRIOR FILING DATE: 2000-09-27
: PRIOR APPLICATION NUMBER: PCT/US01/00666
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00667
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00664
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00669
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00665
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00668
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00663
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00662
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00661
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00670
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: US 60/234,687
: PRIOR FILING DATE: 2000-09-21
: PRIOR APPLICATION NUMBER: US 09/608,408
: PRIOR FILING DATE: 2000-06-30
: PRIOR APPLICATION NUMBER: US 09/774,203
: PRIOR FILING DATE: 2001-01-29
: NUMBER OF SEQ ID NOS: 49117
: SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
: SEQ ID NO 1653
: LENGTH: 454
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: OTHER INFORMATION: MAP TO AP000163.1
: OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.4
: OTHER INFORMATION: EXPRESSED IN HL100, SIGNAL = 1.4
: OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.7
: OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.92
: OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.1
: OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.8
: OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.5
: OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.5
: OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1
: OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.95
```

US-09-864-761-1653

Alignment Scores:

Pred. No.: 1,08e-56 Length: 454
 Score: 650.00 Matches: 114
 Percent Similarity: 98.31% Conservative: 2
 Best Local Similarity: 96.61% Mismatches: 2
 Query Match: 26.60% Indels: 0
 DB: 10 Gaps: 0

US-09-902-772-2 (1-451) x US-09-864-761-1653 (1-454)

QY 334 TyrAspIysLeuSerArgAlaLeuArgTyrTyrAspIysAsnIleMetThrLysVal 353
 Db 452 TACATAAGTCACGCGCGCCCTCGTACTACTATGACAGAATCATGACCAAGGTC 393
 QY 354 HisGlyAsnArgTyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnPro 373
 Db 392 CATGGGAAGCGCTACGCTTACAGTTCACATTCACAGGGATCGCCACAGCCCTCCAGCC 333
 QY 374 HisProGlnSerSerMetTyrLysTyrProSerAspLeuProTyrMetSerSerTyr 393
 Db 332 CACCCCGCGGAGTCATCTCTGTACAGACCCCTCAGACCTCCGTCATGGGCTCTAT 273
 QY 394 HisAlaHisProGlnLysMetAsnPheValAlaProHisProAlaLeuProValThr 413
 Db 272 CACCCCAACCCACAGAGATGACTTGTGGCGCCCACTCCAGCCCTCCCGTGACA 213
 QY 414 SerSerSerPhePheAlaAlaProAsnProTyrThrAsnSerProThrGlyLysLeuThr 433
 Db 212 TCTTCAGATTCTTTTGGCGCCCAACCATCTAGTAATGACCAACGCGGGATATATAC 153
 QY 434 ProAsnThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr 451
 Db 152 CCAACACTAGGCTCCCAACAGCAGCATATGCTTCTCATCTGGGACTTACTAC 99

RESULT 6

US-09-850-799-1
 ; Sequence 1, Application US/09850799
 ; Patent No. US20020090647A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Deneris, Evan S.
 ; APPLICANT: Eyodoro, Dmitry V.
 ; APPLICANT: Hendricks, Timothy J.
 ; TITLE OF INVENTION: Reagents and Methods for the Screening of Compounds
 ; TITLE OF INVENTION: Useful in the Treatment of Neurological Diseases
 ; FILE REFERENCE: CASH-03828
 ; CURRENT APPLICATION NUMBER: US/09/850,799
 ; PRIOR APPLICATION NUMBER: 2001-05-08
 ; PRIOR FILING DATE: EARLIER APPLICATION NUMBER: 09/360,779
 ; NUMBER OF SEQ ID NOS: 13
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 1
 ; LENGTH: 1752
 ; TYPE: DNA
 ; ORGANISM: Rattus norvegicus
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (112)..(1131)
 ; US-09-850-799-1

Alignment Scores:

Pred. No.: 6.62e-42 Length: 1752
 Score: 511.00 Matches: 123
 Percent Similarity: 50.69% Conservative: 23
 Best Local Similarity: 42.71% Mismatches: 63
 Query Match: 20.91% Indels: 79
 DB: 10 Gaps: 10

US-09-902-772-2 (1-451) x US-09-850-799-1 (1-1752)

QY 209 ValTyrProGlnAlaThrGlnArgIleThrThrArgProAspLeuProTyrGlnGlnAla 228

Db 369 GTGCCCGCCCGTGGCGGTCGCCCGCCCATCTCCACCGCCCA----- 407
 QY 229 ArgArgSerAlaIatPrpThrSerHisSerHisProThrGlnSerLysAlaThrGlnProSer 248
 Db 408 -----GTCCCGCGCAGCATGAGACGAGCGGACCTCCAGCCCTCT 452
 QY 249 SerSerThrValProLysThrGlnAspGlnArg-----Pro 260
 Db 453 GATCAACATGTACCTACCAAGATCCCGTCGAGATGCTTTTAAAGAAAGAGAGAGCC 512
 QY 261 GlnLeuAspProTyrGlnIleLeuGlyProThrSerSerArgLeuAlaAsnProGlySe 280
 Db 513 GAGCTGGGGCGC-----CTGAGCCCTGGGTACAGAAA-----GGCAG 551
 QY 280 rGlyGlnIleGlnLeuTyrPglInPheLeuGlnLeuLeuSerAspSerAsnSerAs 300
 Db 552 CGGGCAGATCCAGTGTGTGAGTTTCTAGAGTGTGGCAGACCGCGGAGACGCGG 611
 QY 300 nCysIleThrTyrPglGlyThrAsnGlyLupHelyMetThrAspProAspGluValAl 320
 Db 612 CTGCATCCGCGTGGGAGGGCGGCGCAGGGAGTTCAAGCTCACCGACCCGAGAGTGGC 671
 QY 320 aArgArgTyrPglGlyLupHelySerLysProAsnMetAsnTyrAspLysLeuSerArgAl 340
 Db 672 GCGAGCTGGGGCGGCGCAGAGCAGACCCAAATATGAACATGACAGCAAGTAACTGAGAC 731
 QY 340 aLeuArgTyrTyrTyrAspLysAsnIleMetThrLysValHisGlyLysArgTyrAlaTy 360
 Db 732 ACTGCGTACTACTACAGCAAAACATCATGTGCAAGGTGACAGCGCAAGCGCTACGCCA 791
 QY 360 rLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnProHisPro----- 375
 Db 792 CGCGTTTACTTCCAGGGCTGGCAGAGCTTCCACAGCACACCGCGGAGCGCCAGCG 851
 QY 376 -----ProGlnSerSerMetTyrLysTyrPr 384
 Db 852 CGCCGCTCGCCCGCGCGCAGCGGCGAGCGCGCCGAGATGCGGCATTCACACTCC 911
 QY 384 oSerAspLeu-----ProTyrMetSerSerTyrHisAlaHisProGlnLys 399
 Db 912 GCGTGTCTGCTCCACTGCTCCCTCCCGGCTCTCC-----AA 950
 QY 399 smetAsnPheValAlaProHisProProAlaLeuProValThrSerSerSerPheAla 419
 Db 951 ACTCAACTTATGAGCAAGCTCGGCGGCGCTGGCGCCCTGCTCTCT----- 999
 QY 419 aAlaProAsnProTyrTyrPasnSerPro-----ThrGlyG1 431
 Db 1000 -----TACTGCGCTGATCCCAAGCCAGCCGCGTGGCGCGGCGGCGGCTGC 1046
 QY 431 yIleTyrProAsnThrArgLeuPro-----AlaAlaHisMetPr 444
 Db 1047 GCTTACCCAAACCCCGGGCTGTGACAGCCCTCCGCGGCTTTGGCGGGTGGCGCGCG 1106
 QY 444 oSerHisLeuGlyThrTyrTyr 451
 Db 1107 TTCGCACCTGGGGGCTCATTA 1128

RESULT 7

US-09-920-300A-1716
 ; Sequence 1716, Application US/09920300A
 ; Patent No. US20020136728A1
 ; GENERAL INFORMATION:
 ; APPLICANT: King, Gordon E.
 ; APPLICANT: Meagher, Madeline Joy
 ; APPLICANT: Xu, Jiangchun
 ; APPLICANT: Secrist, Heather
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
 ; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
 ; FILE REFERENCE: 210121.547
 ; CURRENT APPLICATION NUMBER: US/09/920,300A
 ; CURRENT FILING DATE: 2001-07-31

```

; NUMBER OF SEQ ID NOS: 1789
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1716
; LENGTH: 2188
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-920-300A-1716

Alignment Scores:
Pred. No.: 4.21e-35      Length: 2188
Score: 444.50           Matches: 125
Percent Similarity: 41.28%      Conservative: 43
Best Local Similarity: 30.71%   Mismatches: 105
Query Match: 18.19%           Indels: 135
DB: 10                      Gaps: 13

US-09-902-772-2 (1-451) x US-09-920-300A-1716 (1-2188)
QY 122 GluArgValIleValProAlaAspProThrLeuTrpSerThrAspHisValArgGln 141
DB 491 CACGGGCGCTGGGCAATTCCAAAGAACCCCTGGCTGTGGAGAGCAACAGATAGCCAG 550
QY 142 TrpLeuGluTrpAlaValAlaLysGluTyrGlyLeuProAspValAspIleLeuLeuPheGln 161
DB 551 TGGCTTCTCTGGGCCACCAATGAGTTCAGCTGTGAGCTGATCTGCAGAGGTTTC--- 607
QY 162 AsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArgLeuThrPro 181
DB 608 GGCATGAATGGCCAGATGCTGTGTAACCTTGCAAGAGAACCTTGCTGAGCTGGCACCT 667
QY 182 SerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGlnArgGlyAlaThr 201
DB 668 GACTTTGTGGGTGACATTCCTCTGGAAACATCTGGACCAATGATCAAAAGAAACCAA--- 724
QY 202 PheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgIleThrThrArgPro 221
DB 725 -----GAAAAGACAGAGATCAATATGAAAGAAATTCACAACTCAGCTCCGTTCTT 775
QY 222 ----- 228
DB 776 CATGGATTAAACAGACATATAGGTTTGGCAGACAGAGCGCCCTGTGGAATGACAG 835
QY 228 aatgarg-----SerIleTrpThrSer-----HisSerHisProth 240
DB 836 ACACAGAAATTTACCCCAAAGGGGCGCTCTGGACAGATGTGTCCGGCCTCCACACCCAGC 895
QY 240 rglinsersalatrhgln-----Proserse 249
DB 896 GTCACCTAGCTGTGAGCAGAGGTTTCAAGATGTTCGCCAAGTCTGGCTCAGCTCCGTACAG 955
QY 249 rserthr----- 251
DB 956 GTCACCTAGCTGTCTGTCAAGTCAAGACTTCCAGGACGAACTTGAAATTTGCTCACCACAC 1015
QY 252 -----ValProLysThr-----GluAs 257
DB 1016 AATTCTGGAGACTCCCAAGAGCAGACTCCCTGAGAGAGGTGGAGACAGCTTGCAGAGAG 1075
QY 257 pglinaTgProGlnLeuAspProTyrGlnIleLeuGlyPro----- 270
DB 1076 TCAGACTCCCTCTCCAGTCCGTGGAACAGCAGTGTCTGTCTGTGATGTGCAACGGGTTT 1135
QY 270 ----- 270
DB 1136 CCTTCCTTGAGAGCTTCGAAGATGACTGACGACGACTCTCTGTGCTCAATATAGCCACAC 1195
QY 271 -----ThrSerSerArgLeuAlaAsnPro----- 278
DB 1196 AAGTCTTTCAAGATTATCATCCAAAGAGAG-AGTGACCCAGTGGAGCAAGGCAAAACCAAGT 1254
QY 279 -----glyserGlyGlnIleGlnLeuThrPheGln 288
DB 1255 TTTACTGACAGCTGTGTCTGGCCGGCTTTCACAGAGAGTGGACCACTATTTCAGCTGTGGCAGTT 1314

```

```

QY 288 eLeuLeuGluLeuLeuSerAspSerSerAsnSerAsnCysIleThrTrpGluGlyThrAs 308
DB 1315 TCTCTGGAGACTGCTATCAAGAACAAATCCCTGCAAGTATCATCATGAGCTGAGACGCG 1374
QY 308 nGlyGluPheLysMetThrAspProAspGluValAlaArgArgTyrGlyLysGlySer 328
DB 1375 ATGGAGATTAACTGCTGCCGCCCAATGAGTGGGCCCGCGGGGAGAAAGAGAAAAA 1434
QY 328 rLysProAsnMetAsnTyrAspLysLeuSerArgAlaLeuArgTyrTyrTyrAspLysAs 348
DB 1435 TAAGCCCAAGATATACATACAGAGAGAGTGAACCGCGGCTTACCTACTATATACCAAGAA 1494
QY 348 nIleMetThrLysValHisGlyLysArgTyrAlaTyrLysPhe----- 362
DB 1495 CATCATTCACAAAGACGTGGGGAAGCGCTACGTACCGCTGTGTGCGACCTCCAGAA 1554
QY 363 -----AspPheHisGlyIleAlaGlnIleLeuGlnProHisPr 375
DB 1555 CTTCGTGGGGTTACGCCCGGAGAACTGCACGCCATCTGTG---GGCGTCCAGCCGACAC 1611
QY 375 oProGluSerSerMetTyrLysTyrProSerAspLeuPro---TyrMetSerSerTyrH 394
DB 1612 GGAGAGACTGAGTTCGCCGGGAGACCACTGAGCCGCCGCCAGCTCGTGGAGTGGGA 1671
QY 394 salahisProGlnLysMet 400
DB 1672 AGCCCATCTGACCACTG 1690

RESULT 8
US-10-033-528-1716
; Sequence 1716, Application US/10033528
; Patent No. US20020131971A1
; GENERAL INFORMATION:
; APPLICANT: King, Gordon E.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Xu, Jianshun
; TITLE OF INVENTION: SECRETIST, HEATHER
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.547C1
; CURRENT APPLICATION NUMBER: US/10/033,528
; CURRENT FILING DATE: 2001-12-26
; NUMBER OF SEQ ID NOS: 1896
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1716
; LENGTH: 2188
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-033-528-1716

Alignment Scores:
Pred. No.: 4.21e-35      Length: 2188
Score: 444.50           Matches: 125
Percent Similarity: 41.28%      Conservative: 43
Best Local Similarity: 30.71%   Mismatches: 105
Query Match: 18.19%           Indels: 135
DB: 12                      Gaps: 13

US-09-902-772-2 (1-451) x US-10-033-528-1716 (1-2188)
QY 122 GluArgValIleValProAlaAspProThrLeuTrpSerThrAspHisValArgGln 141
DB 491 CACGGGCGCTGGGCAATTCCAAAGAACCCCTGGCTGTGGAGAGCAACAGATAGCCAG 550
QY 142 TrpLeuGluTrpAlaValAlaLysGluTyrGlyLeuProAspValAspIleLeuLeuPheGln 161
DB 551 TGGCTTCTCTGGGCCACCAATGAGTTCAGCTGTGAGCTGATCTGCAGAGGTTTC--- 607
QY 162 AsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArgLeuThrPro 181
DB 608 GGCATGAATGGCCAGATGCTGTGTAACCTTGCAAGAGAACCTTGCTGAGCTGGCACCT 667

```

```
QY 182 SerTyrAsnAlaAspIleuLeuSerHisLeuHisTyrLeuArgGluArgGlyAlaThr 201
    ::::::::::::::::::::
Db 668 GACTTTGGGTGACATCTCTGGAGACATCTGGAGCAAAATGATCAAGAAACCA--- 724
QY 202 PheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgIleThrThrArgPro 221
    ::::::::::::::::::::
Db 725 -----GAAAAGACAGAAAGATCAATATGAAAGAAAATTACACACCTCCACCTCGTTCT 775
QY 222 -----AspLeuProTyr-GluGlnAl 228
    ::::::::::::::::::::
Db 776 CATTGATTAACAGCAATACATAGTTTGGACAGAGAGGCGCCCTATGAAATGACAG 835
QY 228 aArgArg-----SerAlaTyrPheSer-----HisSerHisProThr 240
    ::::::::::::::::::::
Db 836 ACACAGAAATTAACCCCAAGGCGCTCTGACAGCATGTGTCGGCCCTCCACACCCAGC 895
QY 240 rGlnSerLysAlaThrGln-----ProSerse 249
    ::::::::::::::::::::
Db 896 GTACTCAGCTCTGACAGAGAGTTTCAGATGTTCCTCCCAAGTCCTCGCTCAGCTCCGTCAGC 955
QY 249 rSerThr----- 251
    ::::::::::::::::::::
Db 956 GTACACTACTGCTCTGTTCAGTCAAGACTTCCAGGACAGCAACTGAAATTGCTCACCAC 1015
QY 252 -----ValProLysThr-----GluAs 257
    ::::::::::::::::::::
Db 1016 AATTCTGGAGACTCCCAACAGCAAGACTCCCTGAGAACGGTGGGACAGCTTCGAGAGC 1075
QY 257 pGlnArgProGlnLeuAspProTyrGlnIleuGlyPro----- 270
    ::::::::::::::::::::
Db 1076 TCGACACTCCCTCCCTCCAGCTCTGGAGACAGCAAGTCTGCTTGTGATGTGCACAGGGGTT 1135
QY 270 ----- 270
    ::::::::::::::::::::
Db 1136 CCTTCCTCGAGAGCTTCGAAGATGACTGCAGCCAGTCTCTGCTCCTCATTAAGCCAAAC 1195
QY 271 -----TherSerArgLeuAlaAsnPro----- 278
    ::::::::::::::::::::
Db 1196 ATGTCTTCAAGATATACATCCAGAGAGAG-AGTGACCCAGTGAGCAAGGCAAGCAACCACT 1254
QY 279 -----GlySerGlyGlnIleGlnLeuThrProGln 288
    ::::::::::::::::::::
Db 1255 TATACCTGACAGCTGTGCTGGCCGGCTTCACAGAGAGTGGACCTATTCAGCTGTGGAGTT 1314
QY 288 eLeuLeuGlnLeuLeuSerAspSerSerAsnSerAsnGlyIleThrTrpGlnGlyThrAs 308
    ::::::::::::::::::::
Db 1315 TCTCTGGAGCTGTATCAGCAAAATCTCGCACTCATTCATCATGCTGAGCTGAGACGG 1374
QY 308 nGlyGluPheLysMetThrAspProAspGluValAlaArgArgTrpGlnGlyThrGlyse 328
    ::::::::::::::::::::
Db 1375 ATGGGAGTTTAAGCTCGCCAGACCCGATGAGTGGCCGGCGGGGGAAGAGAGAAAA 1434
QY 328 rLysProAsnMetAsnTyrAspLysLysSerArgAlaLeuArgTyrTyrTyrAspLysAs 348
    ::::::::::::::::::::
Db 1435 TAAAGCCAAATGATCACTACAGAGAGAGTGAAGCCGGGCTTACGCTACTATTAACAGAA 1494
QY 348 nIleMetThrLysValHisGlyLysArgTyrAlaTyrLysPhe----- 362
    ::::::::::::::::::::
Db 1495 CAGCATCAACAAGAGCTGGGGAAGCGCTAGCTGACGGTTCGTGTCGACGCTCCAGAA 1554
QY 363 -----AspPheHisGlyIleAlaGlnAlaLeuGlnProHisPr 375
    ::::::::::::::::::::
Db 1555 CTGCTGGGGGTTCAGCCGCCAGAGAACTGACAGCCATCTG---GGCCTCCAGCCGACAC 1611
QY 375 oProGlnSerSerMetLysTyrLysTyrProSerAspLeuPro---TyrMetSerSerTyrH 394
    ::::::::::::::::::::
Db 1612 GGAAGACTGAGGTGGCGGAGACCACTTGAAGCCGCCAGGCTCTGAGCATGAGTGGGA 1671
QY 394 sAlaHisProGlnLysMet 400
    ::::::::::::::::::::
Db 1672 AGCCATCTGTACACCACTG 1690
RESULT 9
```

```
US-09-920-300A-1693
; Sequence 1693, Application US/09920300A
; Patent No. US20020136728A1
; GENERAL INFORMATION:
; APPLICANT: King, Gordon E.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Xu, Jiangchun
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.547
; CURRENT APPLICATION NUMBER: US/09/920, 300A
; CURRENT FILING DATE: 2001-07-31
; NUMBER OF SEQ ID NOS: 1789
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1693
; LENGTH: 2268
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-920-300A-1693

Alignment Scores:
Pred. No.: 4 43e-35 Length: 2268
Score: 444.50 Matches: 125
Percent Similarity: 41.28% Conservative: 43
Best Local Similarity: 30.71% Mismatches: 105
Query Match: 18.19% Indels: 135
DB: 10 Gaps: 13

US-09-902-772-2 (1-451) x US-09-920-300A-1693 (1-2268)
QY 122 GluArgArgValIleValProAlaAspProThrLeuTrpSerThrAspHisValArgGln 141
    ::::::::::::::::::::
Db 571 CAGCGGGCGCTGGGCAATTCCAAGAACCCCTGGCTGTGAGATGAGCAACAGTGTGGCAG 630
QY 142 TrpLeuGluTrpAlaValLysGluTyrGlyLeuProAspValAspIleLeuLeuPheGln 161
    ::::::::::::::::::::
Db 631 TGGCTTCTGCGGCGCACCAATGAGACTTCAGTCTGTGAACTGGAATCTGCAGAGGTTG--- 687
QY 162 AsnIleAspGlyLysGlnLeuCysLysMetThrLysAspAspPheGlnArgLeuThrPro 181
    ::::::::::::::::::::
Db 688 GGCATGAAATGGCCAGATGCTGTGTACCTTGGACAGGACCTTTTCTGAGACTGGACCT 747
QY 182 SerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGluArgGlyAlaThr 201
    ::::::::::::::::::::
Db 748 GACTTTGGGTGACATCTCTGGAGACATCTGGAGCAAAATGATCAAGAAACCA--- 804
QY 202 PheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgIleThrThrArgPro 221
    ::::::::::::::::::::
Db 805 -----GAAAAGACAGAAAGATCAATATGAAAGAAAATTACACACCTCCACCTCGTTCT 855
QY 222 -----AspLeuProTyr-GluGlnAl 228
    ::::::::::::::::::::
Db 856 CATTGATTAACAGCAATACATAGTTTGGACAGAGAGGCGCCCTATGAAATGACAG 915
QY 228 aArgArg-----SerAlaTyrPheSer-----HisSerHisProThr 240
    ::::::::::::::::::::
Db 916 ACACAGAAATTAACCCCAAGGCGCTCTGACAGCATGTGTCGGCCCTCCACACCCAGC 975
QY 240 rGlnSerLysAlaThrGln-----ProSerse 249
    ::::::::::::::::::::
Db 976 GTACTCAGCTCTGACAGAGAGTTTCAGATGTTCCTCCCAAGTCCTCGCTCAGCTCCGTCAGC 1035
QY 249 rSerThr----- 251
    ::::::::::::::::::::
Db 1036 GTACACTACTGCTCTGTTCAGTCAAGACTTCCAGGACAGCAACTGAAATTGCTCACCAC 1095
QY 252 -----ValProLysThr-----GluAs 257
    ::::::::::::::::::::
Db 1096 AATTCTGGAGACTCCCAAGACAGCAAGACTCCCTGAGAACGGTGGGACAGCTTCGAGAGC 1155
QY 257 pGlnArgProGlnLeuAspProTyrGlnIleuGlyPro----- 270
    ::::::::::::::::::::
```

```

Db 1156 TCAGACTCCCTCCTCCAGTCCTGGAAACAGCACTCCTCCTGCTGATGTCAGACGGGTT 1215
Qy 270 -----
Db 1216 CCTTCCTGGAGAGCTTCGAAGATGACTGCAGCCAGCTCTCTGCTCATTAAGCCAAAC 1275
Qy 271 -----ThSerSerArgLeuAlaAsnPro----- 278
Db 1276 AAGTCTTTCAAGGATTACATCCAAAGAGG-AGTGACCCAGTGGACCAAGCAACCACT 1334
Qy 279 -----gluSerGluInIleGlnLeuTrpGlnPh 288
Db 1335 TATACCTGCAGCTGTGCTGGCCGCTTCACAGGAAGTGGACCTATTCAGCTGGGCAGTT 1394
Qy 288 eleuLeuGluLeuLeuSerAspSerSerAsnSerAsnCysIleTrpTrpGluIleTrpAs 308
Db 1395 TCTCCCTGGAGCTGCTATTCAGCAAAATCTGCAGTATTCATCATCAGCTGGACGAGC 1454
Qy 308 nglYgluPhelyMetThrAspProaspGluValAlaArgArgTrpGluIleGluysSe 328
Db 1455 ATGGAGATTAAAGCTGCGCCGACCCGATGAGGTGGCCCGGGTGGGAAAGGAAAGAAA 1514
Qy 328 rlySProAsnMetAsnTyfAspLysLeuSerArgAlaLeuArgTyfTyfAspLysAs 348
Db 1515 TAAGCCCAAGATGAAGTACAGAGAGCTGAGCCGGGCTTACCTACTATTACAGCAAGAA 1574
Qy 348 nllMetThrLysValHisGlyLysArgTyfAlaTyfLysPhe----- 362
Db 1575 CATCATCCACAAGACGCTCGGGAAAGCGCTACGTACCCCTTGCTGTGGACCTCCAGAA 1634
Qy 363 -----AspPheHisGlyIleAlaGlnIleGlnInProHisPr 375
Db 1635 CTTCGCTGGGGTTACGCCCGGAGAGACTGACAGCATCTG--GGCGTCCAGCCGACAC 1691
Qy 375 oProGluSerMetTyfLysTyfProSerAspLeuPro--TyfMetSerSerTyfTrnI 394
Db 1692 GGAGAGCTAGAGTCGCGGGAGACACCTGAGCGGCCCAAGGCTCGTGAGCTAGTGGGA 1751
Qy 394 salaHisProGlnLysMet 400
Db 1752 AGCCCATCTCTGACACGCTG 1770

RESULT 10
US-10-033-528-1693
; Sequence 1693, Application US/10033528
; Patent No. US20020131971A1
; GENERAL INFORMATION:
; APPLICANT: King, Gordon E.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Xu, Jiangchun
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.547C1
; CURRENT APPLICATION NUMBER: US/10/033,528
; CURRENT FILING DATE: 2001-12-26
; NUMBER OF SEQ ID NOS: 1896
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1693
; LENGTH: 2268
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-033-528-1693

Alignment Scores:
Pred. No.: 4,436-35 Length: 2268
Score: 444.50 Matches: 125
Percent Similarity: 41.288 Conservative: 43
Best Local Similarity: 30.718 Mismatches: 105
Query Match: 18.198 Indels: 135
DB: 12 Gaps: 13
US-09-902-772-2 (1-451) x US-10-033-528-1693 (1-2268)

```

```

Qy 122 GluArgValIleValProAlaAspProThLeuTrpSerThrAspHisValArgGln 141
Db 571 CAGCGGCGGCTGGGCAATCCCAAGAAACCCCTGGCTGGAGTGAACAGATATGACAG 630
Qy 142 TrpLeuGluTrpAlaValLysGluTyfGluLeuProAspAlaAspIleLeuLeuPheGln 161
Db 631 TGGCTTCTGGGCGCCAAATGAATTCAGTCTGTGAACCTGAATCTGCAGAGCTTC-- 667
Qy 162 AsnIleAspGluLysGluLeuCysLysMetThrLysAspAspPheGlnArgLeuTrpPro 181
Db 688 GGCATGAATGGCCAGATGCGTGTATACCTTGGCAAGAAACCTTTTGGAGCTGGACCT 747
Qy 182 SerTyfAsnAlaAspIleLeuLeuSerHisLysHisTyfLeuArgGluArgGluAlaTrp 201
Db 748 GACTTGTGGGTGACATCTCTGGGAACATCTGGACCAATATGATCAAGAAACAA-- 804
Qy 202 PheIlePheProAsnTrpSerValTyfTrpGlnAlaIleThrGlnArgIleTrpArgPro 221
Db 805 -----GAAAGACAGAAATCAATATGAAGAAATTCACACCTCACCTCCGTCC 855
Qy 222 -----AspLeuProTyf-GluGlnAl 228
Db 856 CATGGATTAAACAGCAATACATTAGCTTTTGGCACAGACAGCGCCCTATGGAAATGCA 915
Qy 228 aArgArg-----SerAlaTrpTrpSer-----HisSerHisProTh 240
Db 916 ACACAGAAATTAACCCCAAGCGGCTCTGGACAGATGTCTCGGCTCCACACCCAG 975
Qy 240 rGlnSerLysAlaThrGln-----ProSerSe 249
Db 976 GTTACTAGCTGTAGAGAGAGGATTTCAGATTTCCSSAAGTCTCGGCTCACCTCCGTAC 1035
Qy 249 rSerThr----- 251
Db 1036 GTACACTACTGCTGTGTCAGTGCAGACTTCCGAGGACACTTGAATTGCTCACCAAC 1095
Qy 252 -----ValProLysThr-----GluAs 257
Db 1096 AATTCTGGAGCTCCCAAGACACAGACTCCCTGAGAAAGTGGGAGACAGCTTGAGAGC 1155
Qy 257 rGlnArgProGlnLeuAspProTyfGlnIleLeuGlyPro----- 270
Db 1156 TCAGACTCCCTCCTCCAGTCTGGAAACAGCACTCCTGCTGATGTCAGACGGGTT 1215
Qy 270 ----- 270
Db 1216 CCTTCCTGGAGAGCTTCGAAGATGACTGCAGCCAGCTCTCTGCTCATTAAGCCAAAC 1275
Qy 271 -----ThSerSerArgLeuAlaAsnPro----- 278
Db 1276 AAGTCTTTCAAGGATTACATCCAAAGAGG-AGTGACCCAGTGGACCAAGCAACCACT 1334
Qy 279 -----gluSerGluInIleGlnLeuTrpGlnPh 288
Db 1335 TATACCTGCAGCTGTGCTGGCCGCTTCACAGGAAGTGGACCTATTCAGCTGGGCAGTT 1394
Qy 288 eleuLeuGluLeuLeuSerAspSerSerAsnSerAsnCysIleTrpTrpGluIleTrpAs 308
Db 1395 TCTCCCTGGAGCTGCTATTCAGCAAAATCTGCAGTATTCATCATCAGCTGGACGAGC 1454
Qy 308 nglYgluPhelyMetThrAspProaspGluValAlaArgArgTrpGluIleGluysSe 328
Db 1455 ATGGAGATTAAAGCTGCGCCGACCCGATGAGGTGGCCCGGGTGGGAAAGGAAAGAAA 1514
Qy 328 rlySProAsnMetAsnTyfAspLysLeuSerArgAlaLeuArgTyfTyfAspLysAs 348
Db 1515 TAAGCCCAAGATGAAGTACAGAGAGCTGAGCCGGGCTTACCTACTATTACAGCAAGAA 1574
Qy 348 nllMetThrLysValHisGlyLysArgTyfAlaTyfLysPhe----- 362
Db 1575 CATCATCCACAAGACGCTCGGGAGAGCGCTACGTGTCACCCCTTGCTGTGGACCTCCAGAA 1634

```


FILE REFERENCE: Aeonica-X-1
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 7427
LENGTH: 553
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AP001426.1
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.1
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.1
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.3
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.76
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.1
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.1
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2
US-09-864-761-7427

Alignment Scores:
Pred. No.: 6,34e-29 Length: 553
Score: 374.00 Matches: 71
Percent Similarity: 100.00% Conservative: 4
Best Local Similarity: 94.67% Mismatches: 0
Query Match: 15.30% Indels: 0
DB: 10 Gaps: 0

US-09-902-772-2 (1-451) x US-09-864-761-7427 (1-553)

QY 5 IllelsglualeuSerValValSerGluaspGlnSerLeuphncIuGysalatyrgly 24
|||||
Db 297 ATCTGAGGACCTTATCATCTGTGAGTGAGGACCACTCTGTGAGTGAGGACGCGA 238
SerProHlsleuAlaIysrhgIumetPrAlaSerSerSerGlnlytyrglyGlnThr 44
:::|||||

Db 237 ACCGACACCTGGCTAAGACAGATGACCGCTCTCTCCAGCAGCTATGACAGACT 178
QY 45 SerLysMetSerProArGValProGlnGlnAspTrpLeuSerGlnProAlaArGVal 64
|||||
Db 177 TCCAAATGAGGCCACGCGCTCCCTCAGCAGGATGGCTGTCTCAACCCGACGAGGTC 118
QY 65 ThlIlelsgmetGluCysasnProAsnGlnValAsnGlySerArg 79
|||||
Db 117 ACCATCAATGGAATGATTAACCTTACGACGATGATGGCTCAAGG 73

RESULT 13
US-09-864-761-21761/C
Sequence 21761, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aeonica-X-1
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 21761
LENGTH: 225
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AP000021.2
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.95
OTHER INFORMATION: NT HIT: M21535.1, EVALUATE 1.00e-113
OTHER INFORMATION: EST HUMAN HIT: AW948996.1, EVALUATE 1.00e-108
OTHER INFORMATION: SWISSPROT HIT: P11308, EVALUATE 3.00e-37
US-09-864-761-21761

Alignment Scores: 3.35e-29 Length: 225
Pred. No.: 371.00 Matches: 67
Score: 100.00% Conservative: 2
Best Local Similarity: 97.10% Mismatches: 0
Query Match: 15.18% Indels: 0
DB: 10 Gaps: 0
US-09-902-772-2 (1-451) x US-09-864-761-21761 (1-225)
QY 129 AlaaspprothrleutserThrasphisValArgIntPleuGluTrrAlaVallys 148
Db 208 TCAGATCCTACGCTATGAGTACAGACCATGTGCGGCACTGGAGTGGCGGTGAAA 149
QY 149 GluTyrGlyLeuProAspValAspIleLeuLeuPheGlnAsnIleAspGlyLysGluLeu 168
Db 148 GAATATGCGCTTCAGAGCTCAACATCTGTATTCAGAAACATCGATGGAGGAAGAACTG 89
QY 169 CysLysMetThrLysAspAspPheGlnArgLeuThrProSerTyrAsnAlaAspIleLeu 188
Db 88 TGCAGATGACCAAGAGGAGCTCCAGAGGCTCACCCACCTACAAACGCCGACATCCTT 29
QY 189 LeuSerHisLeuHisTyrLeuArgGlu 197
Db 28 CTCACATCTCCACTACTCAGAGAG 2
RESULT 14
US-09-864-761-29364/c
; Sequence 29364, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/633,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21

;; PRIOR APPLICATION NUMBER: US 09/608,408
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/774,203
;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 29364
;; LENGTH: 348
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AP001426.1
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.73
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.52
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.45
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.55
;; OTHER INFORMATION: SWISSPROT HIT: P11308, EVALUE 4.00e-38
;; OTHER INFORMATION: NT HIT: M21535.1, EVALUE 1.00e-113
US-09-864-761-29364
Alignment Scores: 6.38e-29 Length: 348
Pred. No.: 371.00 Matches: 67
Score: 100.00% Conservative: 2
Best Local Similarity: 97.10% Mismatches: 0
Query Match: 15.18% Indels: 0
DB: 10 Gaps: 0
US-09-902-772-2 (1-451) x US-09-864-761-29364 (1-348)
QY 129 AlaaspprothrleutserThrasphisValArgIntPleuGluTrrAlaVallys 148
Db 313 TCAGATCCTACGCTATGAGTACAGACCATGTGCGGCACTGGAGTGGCGGTGAAA 254
QY 149 GluTyrGlyLeuProAspValAspIleLeuLeuPheGlnAsnIleAspGlyLysGluLeu 168
Db 253 GAATATGCGCTTCAGAGCTCAACATCTGTATTCAGAAACATCGATGGAGGAAGAACTG 194
QY 169 CysLysMetThrLysAspAspPheGlnArgLeuThrProSerTyrAsnAlaAspIleLeu 188
Db 193 TGCAGATGACCAAGAGGAGCTCCAGAGGCTCACCCACCTACAAACGCCGACATCCTT 134
QY 189 LeuSerHisLeuHisTyrLeuArgGlu 197
Db 133 CTCACATCTCCACTACTCAGAGAG 107
RESULT 15
US-09-864-761-5035/c
; Sequence 5035, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Aecomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/633,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30

```
/ PRIOR APPLICATION NUMBER: PCT/US01/00667
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00664
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00669
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00665
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00668
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00663
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00662
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00661
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00670
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: US 60/234,687
/ PRIOR FILING DATE: 2000-09-21
/ PRIOR APPLICATION NUMBER: US 09/608,408
/ PRIOR FILING DATE: 2000-06-30
/ PRIOR APPLICATION NUMBER: US 09/774,203
/ PRIOR FILING DATE: 2001-01-29
/ NUMBER OF SEQ ID NOS: 49117
/ SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
/ SEQ ID NO 5035
/ LENGTH: 477
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ OTHER INFORMATION: MAP TO AP000021.2
/ OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.95
US-09-864-761-5035
```

```
Alignment Scores:
Pred. No.: 1,02e-28 Length: 477
Score: 371.00 Matches: 67
Percent Similarity: 100.00% Conservative: 2
Best Local Similarity: 97.10% Mismatches: 0
Query Match: 15.18% Indels: 0
DB: 10 Gaps: 0
```

US-09-902-772-2 (1-451) x US-09-864-761-5035 (1-477)

```
QY 129 AAlaSPProThrlEuPrSerThraSPHlSValArgGlnTrpLeuGlnTrpAlaValLys 148
DB 454 TCAGATCCCTACCGTATGAGTACAGACCATGTGCGGCACTGGCTGAGTGGCGGTGAAA 395
QY 149 GluTyrgLylLeuProAspValAspLleLeuLeuPheGlnAsnLleAspGlyLysGluLeu 168
DB 394 GAATATGGCCCTCCAGNCGTCAACATCTGTATTCCAGAACATCGATGGAGAACTG 335
QY 169 CysLysMetThrLysAspAspPheGlnArgLeuThrProSerTyraSnaLlaAspLleLeu 188
DB 334 TGCAGATGACCAAGAGCAGACTTCCAGAGGCTCACCCCACTACACGCCGACATCTTT 275
QY 189 LeuSerHisLeuHisTyrlLeuArgGlu 197
DB 274 CTCTCACATCTCCACATCCTCAGAGAG 248
```

Search completed: November 9, 2002, 21:28:37
Job time : 73.0527 secs

JOURNAL Unpublished (2002)
 COMMENT Contact: Buerstedde JM
 Cellular Immunology
 Heinrich-Pette-Institute
 Martinstr. 52, 20251 Hamburg, Germany
 Email: URL: <http://genetics.hpi.uni-hamburg.de/dt40est.html>.

FEATURES
 source 1..827
 /organism="Gallus gallus"
 /db_xref="taxon:9031"
 /clone_id="8ml3r2"
 /clone_id="riken1"
 /cell_type="bursal lymphocyte"
 /dev_stage="2-3 weeks Old"
 /note="CB Inbred strain"

BASE COUNT 233 a 236 c 177 g 181 t

ORIGIN

Alignment Scores:
 Pred. No.: 2,21e-103 Length: 827
 Score: 1118.50 Matches: 211
 Percent Similarity: 78.31% Conservative: 20
 Best Local Similarity: 71.53% Mismatches: 43
 Query Match: 45.77% Indels: 21
 Gaps: 3

US-09-902-772-2 (1-451) x AJ456498 (1-827)

OY 138 HlSValaIrglInThrPleuGlutrrPaIaValIySglutryGlyLeuPProAspValaIle 157
 |||||
 Db 5 CACGTCGCGACAGTGGCTGGAATGGCCATAAAGAGTATGATTAAGAGATTGACAC 64
 |||||

OY 158 LeuLeuPheGlnAsnIleAspGlyLysGluLeuGlyLeuMetThrIysAspPheGln 177
 |||||
 Db 65 ACCCTCTCCAGAAATGATGATGCAAGAGCTCTGCAAAACGAAGATGACTTCTC 124
 |||||

OY 178 ArgLeuThrProSerTyrAsnIleAspIleLeuLeuSerHisIleuHisTyrLeuArgGlu 197
 |||||
 Db 125 CGAACCACTCCCTCTACACACAGAGATGTTCTTCTCAGTCACTTACCTCAGGGA 184
 |||||

OY 198 ArgGlyAlaThrPheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgIle 217
 |||||
 Db 185 AGTAGCTCACTGCTGCTAC--AATACTCATCCACACAGAACTCTCAGCTCT 241
 |||||

OY 218 ThrThrArgProAspLeuProTyrGluGlnAlaArgArgSerAlaThrSerHisSer 237
 |||||
 Db 242 GGCACCAAGAAAGTCTCTCTGTT----- 265
 |||||

OY 238 HlSProThrGlnSerIySAlaThrGlnProSerSerThrValProLysThrGluAsp 257
 |||||
 Db 266 -----GCAGGACACAAAACGTGACACAGACA-----ACAGAACAG 301
 |||||

OY 258 GlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSerArgLeuAlaAsn 277
 |||||
 Db 302 CAACGGCCCTCAGACAGATCCATCAATCCGCGGCCACAGTAGTGTCTGTGCCAAT 361
 |||||

OY 278 ProGlySerGlyGlnIleGlnLeuThrGlnPheLeuLeuGluLeuSerAspSer 297
 |||||
 Db 362 CCTGGAGAGGGCAGATACACTATGGCAGTCTCTGAGTGTGTCGACAGATTCC 421
 |||||

OY 298 AsnSerAsnCysIleThrTrpGluGlnIySThrAsnGlyLysPheLysMetHisAspProAsp 317
 |||||
 Db 422 AACGCGACGTGATTCATCATGGAGAGGCCAATGGGAATTCAGATGACAGCCAGAT 481
 |||||

OY 318 GluValAlaArgArgTrpGlyLysSerIySProAsnMetAsnTyrAspLysLeu 337
 |||||
 Db 482 GAAGTGGACGCGCTGGGAGAGACCAAGCAAGCCCAACATGATATGACAAAGCTG 541
 |||||

OY 338 SerArgAlaLeuArgTyrTyrIySAsnIleMetThrLysValHisGlyLysArg 357
 |||||
 Db 542 AACCGAGCCCTTCGATCTACTATGATTAAGAACATTATGACCAAGATGATGCAAAAG 601
 |||||

OY 358 TyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnProHisProGlu 377
 |||||

Db 602 TATGCTACAAATTGACTTTCATGGCATTGGCCAGGCTCTCCAGCTCATCCACTGAA 661
 |||||

OY 378 SerSerMetLysTyrTyrProSerAspLeuProTyrMetSerSerTyrHisAlaHisPro 397
 |||||

Db 662 TCATCATATGACAGATGATCATGAGATCTCTCTACATGCTCTTACCATGCCACAG 721
 |||||

OY 398 GlnLysMetAsnPheValAlaProHisProProAlaLeuProValThrSerSerPhe 417
 |||||

Db 722 CAGAAAGTGAACCTTGTACCCACACCTTCTTATGCTGTGACATCATCATCAGTTTC 781
 |||||

OY 418 PheAlaAlaProAsnProTyrTrpAsnSerProThrGlyGlyIle 432
 |||||

Db 782 TTCGAGCAGCTTCACCTTATGAGACCTCTGCTGGAGACATT 826
 |||||

RESULT 2
 BM456833
 LOCUS BM456833 1045 bp mRNA linear EST 05-FEB-2002
 DEFINITION AGENCOURT_6404041 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:583753
 5', mRNA sequence.
 ACCESSION BM456833
 VERSION BM456833.1 GI:18505873
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 1045)
 NIH-MGC <http://mhc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLM12347 row: f column: 10
 High quality sequence stop: 669.

FEATURES
 source 1..1045
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_id="IMAGE:583753"
 /clone_id="NIH_MGC_92"
 /tissue_type="embryonal carcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: testis; Vector: pCMV-SPORT6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
 Average insert size 2.5 kb. Library enriched for
 full-length clones and constructed by Life Technologies.
 Note: this is a NIH-MGC Library."

BASE COUNT 300 a 290 c 230 g 225 t

ORIGIN

Alignment Scores:
 Pred. No.: 3.13e-95 Length: 1045
 Score: 1040.00 Matches: 214
 Percent Similarity: 75.24% Conservative: 17
 Best Local Similarity: 69.71% Mismatches: 33
 Query Match: 42.55% Indels: 44
 Gaps: 8

US-09-902-772-2 (1-451) x BM456833 (1-1045)

OY 7 GluAlaLeuSerValValSerGluAspGlnSerLeuPheGluCysAlaTyrGlySerPro 26
 |||||

Db 190 GAAGCTTATCATGTTGTGATGAGGACAGTGTGTTGAGTGTGCTCAGGAGGACCA 249
 |||||

OY 27 HlSLeuAlaLysThrGluMetThrAlaSerSerSerSerGlyTyrGlnThrSerLys 46
 |||||

```

Db 250 CACCTGGCTAAGACAGATGACCGCTCTCTCCAGCCGACTGTGACAGACTTCCAG
Qy 47 MetSerProArgValProGlnGlnAspTrpLeuSerGlnProProAlaArgValThrIle
Db 310 ATGAGCCAGCGCGCTCCAGCAGAGATTGGGTCTCAACCCCAAGCAGGAGTCAATC
Qy 67 LysMetIuCySAsnProAsnGlnValAsnGlySerArgAsnSerProAspPcysSer
Db 370 AAAATGCAAGTAACTTACCTAGCAGTCAAGTGGCTCAAGAACTCTCCGTGATGATGAGT
Qy 87 ValAlaIasGlyGlyLysMetValSerSerSerAspAsnValGlyMetAsnTyrGlySer
Db 430 GTGGCCAAAGCGGGAAAGATGGTGGGAGCCAGACACCGTGGGATGAACTAGCGGAGC
Qy 107 TyrMetGluGluLysHisIleProProProAsnMetThrThrAsnGluArgArgValIle
Db 490 TACATGAGAGAGAAAGACATATGCCACCCCAACATGACACGAAACGAGCGAGTTATC
Qy 127 ValProAlaAspProThrLeuTrpSerThrAspHisValArgGlnTrpLeuGluTrpAla
Db 550 GTGCCACACATCTCTACGCTATGAGTACAGACCATGTCCGCGAGTGGAGTGGGCG
Qy 147 ValIleGluTrpGlyLeuProAspValAspIleLeuLeuPheGlnAsnIleAspGlyLys
Db 610 GTGAAGAATAATGCGCTTCAGACGTCAACATCTGTATTCTCAGAACATCGATGGAGAG
Qy 167 GluLeuCySlysmethrLysAspAspPheGlnArgLeuThrProSerTyrAsnAlaAsp
Db 670 GAACGTGCAAGATGACCAAGAGACGACTCCAGAGGCTCACCCGACTACACAGCCGAC
Qy 187 IleLeuLeuSerHisIleHisIleTyrLeuArgGluArgGlyAlaThrPheIleProAsn
Db 730 ATCTCTCTCTACATCTTCACTACCTCAGAGACTCCTCTTCCACATTTGACTCCAGAT
Qy 207 ThrSerValTyrProGlnAlaThrGlnArgIlePheThrArgPro-----
Db 790 -----GATGTTGATTAAGC-CTTACAAACTCTCCCGGTTAATGCAATGCT 833
Qy 222 -----AspLeuProTyrGluGlnAlaArgArgSerAlaTrpThrSerHisSerHis
Db 834 AGAAACACGATTTACCATATGAGACCCCGGAGATCAGCTGACCGGTCAGACGATCA
Qy 239 ProThrGlnSerLysAlaThrGlnProSerSerSerThrVal-----Pro
Db 894 CCC-----CAAGCCCAAGTCGAAAGCTGTTCAAACATTTCTCCCTTCACAAAGGCGCC
Qy 254 LysThrGluAspGlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSerSer
Db 945 AAAAAGTAA-----AGACCCAGC-----GGCCCTCTCAGGT-- 974
Qy 274 ArgLeuAlaAsnProGlySerGlyGlnIleGlnLeuTrpGlnPheLeuLeuGluLeu
Db 975 -----TAAATATCC-----TAAATCACAATCTT 998
Qy 294 SerAspSerSerAsnSerAsn 300
Db 999 TGGGAACCCCAATATTAAC 1019

RESULT 3
LOCUS BQ233264 880 bp mRNA linear EST 02-MAY-2002
DEFINITION AGENCOUT_7565836 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:6041526
5', mRNA sequence.
ACCESSION BQ233264
VERSION BQ233264.1 GI:20414664
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 880)
AUTHORS NIH-MGC http://mgc.ncl.nih.gov/.

```

```

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LMNL at:
http://image.llnl.gov
Plate: ILM13279 row: h column: 07
High quality sequence stop: 608.
FEATURES
source
1..880
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6041526"
/clone_id="NIH_MGC_92"
/tissue_type="embryonal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: testis; Vector: pCMV-SPORT6; Site:1; Noti;
Average insert size 2.5 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 250 a 231 c 207 g 189 t 3 others
ORIGIN
Alignment Scores:
Pred. No.: 5,98e-93 Length: 880
Score: 1016.50 Matches: 195
Percent Similarity: 84.94% Conservative: 8
Best Local Similarity: 81.59% Mismatches: 7
Query Match: 41.59% Indels: 29
DB: 14 Gaps: 2
US-09-902-772-2 (1-451) x BQ233264 (1-880)
Qy 1 MetLaserThrIleLysGluAlaLeuSerValValSerGluAspGlnSerLeuPheGlu 20
Db 74 ATGGCCAGCAGCATTAAGAAAGGCTTTACGATTGTAGTGAGGACCAAGCTGTTTGG 133
Qy 21 CysAlaTyrGlySerProHisIleuAlaLysThrGluMetThrAlaSerSerSerGlu 40
Db 134 TGTGCTTACGGAACGCCACACCTGCTAAGACAGATGACCGGCTCTCTCCAGGAGC 193
Qy 41 TyrGlyGlnThrSerLysMetSerProArgValProGlnGlnAspTrpLeuSerGlnPro 60
Db 194 TATGAGACAGACTTCCAAAGATGAGCCACGCGCTCCAGCAGGATGGCTGTCTCAACCC 253
Qy 61 ProAlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsn 80
Db 254 CCAGCCAGGCTACCATCAAAATGGAATGTAACTTACCGCAGGTAATGGCTCAAGAAC 313
Qy 81 SerProAspAspCysSerValAlaLysGlyLysMetValSerSerSerAspAsnVal 100
Db 314 TCTCCTGATGATGACGATGTGGCCAAAGCGGGAATGGTGGGACAGCCACGAGT 373
Qy 101 GlyMetAsnTyrGlySerTyrMetGluGluLysHisIleProProAsnMetThrThr 120
Db 374 GGGATGAACTACGCGACCTACATGAGAGAGACATGCCACCCCAACATATGACACG 433
Qy 121 AsnGluArgArgValIleValProAlaAspProThrLeuTrpSerThrAspHisValArg 140
Db 434 AACGAGCGCAGAGATTATCGTCCAGCAGATCTTACGCTTATGAGACTACAGACATGTCCG 493
Qy 141 GlnTrpLeuGluTrpAlaValLysGluTyrGlyLeuProAspValAspIleLeuPhe 160
Db 494 CAGTGGCTGAGTGGGCGGAAGATATGCGCTTCCAGACGTCACACATCTTATTC 553
Qy 161 GlnAsnIleAspGlyLysGluLeuCySlysmethrLysAspAspPheGlnArgLeuThr 180

```

```

Db 554 CAGACATGATGGGAGGAGACTGTGTCAGATGACCAAGAGACTTCCAGAGGCTCAC 613
QY 181 ProSerTyraAsnAlaAspIleuLeuSerHisIleuHisTyrIleuArgIuArg----- 198
|||||
Db 614 CCAGCTACAGCCGACATCTCTCTCTCACATCTCCACTACTGAGAGACTCTCT 673
QY 198 ----- 198
Db 674 CCACATTTGACTTCAGATGATGATTAAGCCCTTACAAACTCTCCAGGGGNTATGC 733
QY 199 -----GlyAlaThrPheIlePheProAsn---ThreValTyr 210
Db 734 ATGGCTAGAAACACAGAGGGGGTGCACCTTTTATTTTCCCAATACTTCAGATAT 790
|||||

RESULT 4
BQ212127 1089 bp mRNA linear EST 02-MAY-2002
LOCUS AGENCOURT.7571287 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:6052136
DEFINITION 5', mRNA sequence.
ACCESSION BQ212127
VERSION BQ212127.1 GI:20392058
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 1089)
AUTHORS NIH-MGC http://mhc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM13307 row: b column: 09
High quality sequence stop: 562.
Location/Qualifiers
1. 1089
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6052136"
/clone_lib="NIH_MGC_92"
/tissue_type="embryonal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: testis; Vector: pCMV-Sport6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 2.5 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 292 a 314 c 257 g 225 t 1 others
ORIGIN

Alignment Scores:
Pred. No.: 2.2e-91 Length: 1089
Score: 1002.50 Matches: 207
Percent Similarity: 70.31% Conservative: 18
Best Local Similarity: 64.69% Mismatches: 45
Query Match: 41.02% Indels: 50
DB: 14 Gaps: 5

US-09-902-772-2 (1-451) x BQ212127 (1-1089)
QY 1 MetAlaSerThrTleYsGluAlaIeuSerValValSerGluAspGlnSerIeuPheGlu 20
|||||
Db 76 ATGGCAGACACTATTAAAGACCTTATCAGTTGTGAGAGACAGACGCTTGTAG 135
|||||
QY 21 CysAlaTyrGlySerProHisIeuAlaYsThrGluMetThrAlaSerSerSerGlu 40
|||||

```

```

Db 136 TGTGCTACGGAGACGCCACACTGGCTTAAGACAGAGATGACCGGCTCTCTCCAGCAG 195
QY 41 TyrGlyGlnThrSerYsMetSerProArgValProGlnGlnAspThrIeuSerGlnPro 60
|||||
Db 196 TATGGACAGACTTCCAAAGATGAGCCACAGGCTCCCTCAGAGATTGGCTGTCAACC 255
QY 61 ProAlaArgValThrIleYsMetGluCysAsnProAsnGlnValAsnGlySerArgAsn 80
|||||
Db 256 CCAGCCAGGGTCCACATCAAAAAGATGTAACTCAGCCAGGTGATGGCTCAAGAGAC 315
QY 81 SerProAspAspCysSerValAlaIuYsGlyGlyYsMetValSerSerSerAspAsnVal 100
|||||
Db 316 TCTCTGATGAATGACAGTGTGGCCAAAGCGGGAATATGTGGCGACCCAGACACCTGT 375
QY 101 GlyMetAsnTyrGlySerTyrMetGluGluYsHisIleProProAsnMetThrThr 120
|||||
Db 376 GGGATGACATACGAGGAGCTACATGAGGAGAGACATGACACCCCAACATGACCAG 435
QY 121 AsnGluArgArgValIleValProAlaAspProThrIleuThrSerThrAspHisValArg 140
|||||
Db 436 AACGACGCGAGATTATCTGTCAGCAGATCTACGCTATGGAGTACAGACCATGTGCGG 495
QY 141 GlnThrIeuGluThrAlaValIleYsGluTyrGlyIleuProAspValIleIeuIeuPhe 160
|||||
Db 496 CAGTGGCTGGAGTGGCGGGGAGAGATATGGCTTCCAGACTCAACATCTTGTATTC 555
QY 161 GlnAsnIleAspGlyYsGluIeuCysYsMetThrIleYsAspAspPheGlnArgIeuThr 180
|||||
Db 556 CAGAACATCATGGAAGGAGACGTGCACAGATATACCAACAGAGACACTCCAGAGGCTCAC 615
QY 181 -ProSerTyraAsnAlaAspIleuLeuSerHisIleuHisTyrIleuArgIuArg----- 198
|||||
Db 616 CCCACATCATAGGCGGACATCTCTCTCACATCTCCACTACCTGAGAGACTCTCT 675
QY 198 ----- 198
Db 676 TCCACATTTGACTACATGATGTTGATTAAGCCCTTAAACTCTCCAGGGGTAATGC 735
QY 199 -----GlyAlaThrPheIlePheProAsnThr-----SerValTyr 210
Db 736 ATGGCTAAACCCAGAGGGGGGCGACACTTTTATTTTCCCAATTCATTGAGAAATTCCT 795
QY 210 rProGlnAlaThrGlnArgIleThrThrArg-ProAspIeuProTyrGluGlnAlaArg 230
Db 796 GGAAGCTACGCCCAAGAAATTACACCTAGGGCCAAATTACCCCTATGAGACCCGCC 855
QY 230 rg-----SerAlaThrThr-SerHisSerHisProThrGlnSerYsAlaThrGlnPro 247
Db 856 GGAAGATCACGCCCTGTGACCGGCTCATGGCCACCC----- 891
QY 248 SerSerSerThrValProYsThrGlnAspGlnArgProGlnIleuAspProTyrGlnIle 267
Db 892 -----CCGACGCGCGAGGTGCGAAGGAGGCTGGCCGCCCATTCCT----- 933
QY 268 IeuGlyProThrSerSerArgIeuAlaAsnProGlySerGlyGlnIleGlnIeu 285
Db 934 ---TCCCGCGGGGCCCAACCTGAAAGAACACAGCGGTCCCTCGGTAAATC 984

RESULT 5
LOCUS BG388025 715 bp mRNA linear EST 12-MAR-2001
DEFINITION 602412867f1 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:4521233 5',
mRNA sequence.
ACCESSION BG388025
VERSION BG388025.1 GI:13281471
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 715)
AUTHORS NIH-MGC http://mhc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

```

JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 Tissue Procurement: ATCC

CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
 Plate: LLM10420 row: f column: 18
 High quality sequence stop: 693.
 Location/Qualifiers

FEATURES

source

1. 715
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4521233"
 /clone_1lb="NIH_MGC_92"
 /tissue_type="embryonal carcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: testis; Vector: PCMV-SPORT6; Site 1: NotI;
 Site 2: SalI; Cloned unidirectionally; oligo-dt primed.
 Average insert size 2.5 kb. Library enriched for
 full-length clones and constructed by Life Technologies.
 Note: this is a NIH_MGC Library."

BASE COUNT 206 a 188 c 172 g 149 t
 ORIGIN

Alignment Scores:

Pred. No.: 2.64e-91 Length: 715
 Score: 999.00 Matches: 186
 Percent Similarity: 97.46% Conservative: 6
 Best Local Similarity: 94.42% Mismatches: 5
 Query Match: 40.88% Indels: 0
 DB: 12 Gaps: 0

US-09-902-772-2 (1-451) x BG386025 (1-715)

QY 1 MetAlaSerThrIleTylGluAlaLeuSerValSerGluAspGlnSerLeuPheGlu 20
 Db 94 ATGGCCAGCACTATTAAAGAAAGCCTTATCACTGTGAGTGAGGACCACTGCTTTTGAG 153
 QY 21 CysAlaTyrGlySerProHisLeuAlaLysThrGluMetThrAlaSerSerSerGlu 40
 Db 154 TGTGCTCAAGGAAAGCCACACCTGCTAGACAGATGACCGGCTCCTCCACAGCAG 213
 QY 41 TyrGlyLinhTrSerLysMetSerProArgValProGlnGlnAspTrpLeuSerGlnPro 60
 Db 214 TATGAGACAGACTTCCAAGATGAGCCACGCGCTCCACAGGAGATGGCTGCTCAACCC 273
 QY 61 ProAlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsn 80
 Db 274 CCAGCCAGGGGCAACCAATAATGATGTAACCTTACCCAGGTGATGGCTCAAGGAG 333
 QY 81 SerProAspAspCysSerValAlaLysGlyLysMetValSerSerSerAspAsnVal 100
 Db 334 TCTCTATGATGATCACTGTCGCAAGCGGGAAGATGGGAGGACCCAGACACCGCT 393
 QY 101 GlyMetAsnTyrGlySerTyrMetGluGluLysHisIleProProProAsnMetThrThr 120
 Db 394 GGGATGAACTACGGCAGCTCATGAGAGAGACATGACCCCAAAACATGACACAG 453
 QY 121 AsnGluArgArgValIleValProAlaAspProThrLeuTrpSerThrAspHisValArg 140
 Db 454 AACGAGGCGAGAGTATCGTCCAGCAGATCTTACGCTATGAGACTACAGACATGTGGG 513
 QY 141 GlnTrpLeuGlnTrpAlaValLysGlyTyrGlyLeuProAspValAspIleLeuLeuPhe 160
 Db 514 CAGTGGCTGAGTGGCGGGAAGAAATATGGCTTCCAGACGTCACATCTGTATTTC 573
 QY 161 GlnAsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArgLeuThr 180
 Db 574 CAGAAATCATGAGGAAGAACTGTGCAAGATGACCAAGGACGACTTCCAGAGGCTCAAC 633

QY 181 ProSerTyrAsnAlaAspIleLeuLeuSerHisIleHisTyrLeuArgGlu 197
 Db 634 CCCAGCTCAACACCGCGACATCTCTCATCATCTCCATATCTACAGAG 684

RESULT 6
 B1558012
 LOCUS
 DEFINITION B1558012 949 bp mRNA linear EST 05-SEP-2001
 603240720P1 NCI_CGAP_Mam4 Mus musculus cDNA clone IMAGE:5293492 5',
 mRNA sequence.

ACCESSION B1558012
 VERSION B1558012.1 GI:15445326
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 Tissue Procurement: Lotmar Hennighausen Ph.D., Priscilla Furth
 Ph.D.

CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
 Plate: LLM11742 row: h column: 05
 High quality sequence stop: 876.
 Location/Qualifiers

FEATURES

source

1. 949
 /organism="Mus musculus"
 /strain="NMRI"
 /db_xref="taxon:10090"
 /clone="IMAGE:5293492"
 /clone_1lb="NCI_CGAP_Mam4"
 /tissue_type="tumor, gross tissue"
 /dev_stage="5 months"
 /lab_host="DH10B"
 /note="Organ: mammary; Vector: PCMV-SPORT6; Site 1: SalI;
 Site 2: NotI; Cloned unidirectionally. Primer: oligo dt.
 Library constructed by Life Technologies. Investigators
 providing samples: Lotmar Hennighausen/Priscilla Furth,
 NIH Reference for transgenic model: Li et al., Cell Growth
 and Differentiation 7, 3-11 (1996)."

BASE COUNT 252 a 288 c 209 g 200 t
 ORIGIN

Alignment Scores:

Pred. No.: 1.08e-89 Length: 949
 Score: 985.00 Matches: 190
 Percent Similarity: 79.71% Conservative: 30
 Best Local Similarity: 68.84% Mismatches: 47
 Query Match: 40.30% Indels: 9
 DB: 13 Gaps: 3

US-09-902-772-2 (1-451) x B1558012 (1-949)

QY 183 TyrAsnAlaAspIleLeuLeuSerHisIleHisTyrLeuArgGluArgGlyAlaThrPhe 202
 Db 2 TACAACACAGAAAGTGTCTGTGCACTCATGTAACCTCAGGAA-----AGTTCACGTG 55
 QY 203 IlePheProAsnTrpSerValTyrProGluAlaThrGlnArgIleThrThrArgProAsp 222
 Db 56 CTGGCCATATACACACACTCCCATACAGACAGCAGTCTCCACAGCAATATGTCAGAGAAC 115
 QY 223 LeuProTyrGluGlnAlaArgSerAlaTrpThrSerHisSerHisProThrGlnSer 242
 Db 116 CCTCTTATGACTGTGTCAGGAGAGAGACATGAAACATATATATGAACTGTGGCTCAAC 175

KEYWORDS	EST.
SOURCE	COW.
ORGANISM	Bos taurus
REFERENCE	Smith, T.P.L., Grose, W.M., Freking, B.A., Roberts, A.J., Stone, R.T., Casas, E., Wray, J.E., White, J., Cho, T., Fahnenkrug, S.C., Bennett, G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A., Chitko-Mckown, C.G., Pettee, G., Holt, I., Karameycheva, S., Liang, F., Quackenbush, J., and Keele, J.W.
AUTHORS	1 (bases 1 to 592)
TITLE	Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle
JOURNAL	Genome Res. 11 (4), 626-630 (2001)
MEDLINE	21180013
COMMENT	Contact: Smith TPL USDA, ARS, US Meat Animal Research Center PO Box 166, Clay Center, NE 68933-0166, USA Tel: 402 762 4366 Fax: 402 762 4390 Email: smitht@mail.marc.usda.gov Single pass sequencing. Bases called and all trimmed with phred v0.980904.e. Vector identified by cross_match with the -minscore 18 and -mismatch 12 options. PCR primers FORWARD: AGGAACACGCTATGACCAT BACKWARD: GTTTCACGTCACGACG Plate: 139 row: D column: 21 Seq primer: ATTACGACGACCTATAG. Location/Qualifiers 1..592 /organism="Bos taurus" /db_xref="taxon:9913" /clone_id="MARC 2BOV" /tissue_type="pooled" /lab_host="DH10B" /note="Vector: pCMV SPORT6; site_1: NotI; site_2: SalI; library made from pooled tissue from testis, thymus, semitendinosus muscle, longissimus muscle, pancreas, adrenal, and endometrium."
FEATURES	
SOURCE	
BASE COUNT	152 a 177 c 159 g 104 t
ORIGIN	
Alignment Scores:	1.71e-89 Length: 592
Pred. No.:	980.00 Matches: 181
Score:	96.89% Conservative: 6
Percent Similarity:	93.78% Mismatches: 0
Best local Similarity:	40.10% Indels: 0
Query Match:	13 Gaps: 0
DB:	
US-09-902-772-2 (1-451) x BM088063 (1-592)	
OY	1 MetAlaSerThrIleIysGluAlaLeuSerValIleSerGluAspGlnSerLeuPheGlu 20
Db	12 ATGGCCACACTATTAAAGAAAGCGTTATCAAGTCGTGACGAGGACCACTGTTTAG 71
OY	21 CysAlaIyrGlySerProHisLeuAlaIyrThrGluMetThrAlaSerSerSerSerGlu 40
Db	72 TGGCGCTAGCGGACGCCGCCACCTGCGCCAAAGACGACATGACCGCTCTCCGCCGAC 131
OY	41 TyrGlyGlnThrSerLysMetSerProArgValProGlnGlnAspTyrLeuSerGlnPro 60
Db	132 TATGACAGACATCCAAATGATGAGCCCGCGCTCCCTCAGCAGGACTGGCTGTCAACC 191
OY	61 ProAlaArgValIleThrIleLysMetGluLysAsnProAsnGlnValAsnGlySerArgAsn 80
Db	192 CCAGCCAGGGTCACCATTAAGATGGAATGTAACTTAACCAAGGTGAACGGCTCAAGGAC 251
OY	81 SerProAspArgCysSerValAlaIleGlyGlyLysMetValSerSerSerAspAsnVal 100
Db	252 TCCCTATATGATGTCAGATGATGCCCAAGGTGGGAAGATGTGTGGCAAGCCGACACCGTC 311

QY	101	GLYMEASNTYRGLYSerTyrMetcLulLysSHSLiePProProAsmethrthr	120
Db	312	GGGATGACACTACAGACGTACATGAGAGAAAGCAATGCCACCCCAACATGACC	371
QY	121	ASnGuAArgValIleValProAlaAspProthrIeutrPserThrAspHisValArg	140
Db	372	AACGAGCCCGCCGTCATGCTCCCTCCATATCCACGCTATGAGAGACACATGTCCGG	431
QY	141	GLNTrpleauGLUTPrAlaValLysGLulurTgLyLeuProAspValAspIleuLeuPhe	160
Db	432	CAGTGGCTGGAGTGGGCGAGTCAAAAGAAACGGGGCTCCGAGCTTGACATCTTATATTC	491
QY	161	GLnSnIleAspGLyLyGLuLeuCysLysMetThrLysAspAspPheGLNArgLeuThr	180
Db	492	CAGATATCGACGGGAAGAGACTGTGCAAGATGACAAAGATGACTTCACAGGCTCAC	551
QY	181	ProSerTyraSnAlaAspIleuLeuSerHisLeuHis	193
Db	552	CCGAGCTACACGCTGACATCTTCTGTGCGACCTCCAC	590
RESULT 10			
LOCUS	BG390291	915 bp	MRNA
DEFINITION	60241625F01 NIH_MGC_92 Homo sapiens CDNA clone IMAGE:4524852.5,	linear	EST 12-MAR-2
ACCESSION	BG390291		
VERSION	BG390291.1	GI:13283739	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	1 (bases 1 to 915)		
JOURNAL	NIH-MGC http://mgc.nci.nih.gov/ .		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: cgapsb-remail.nih.gov		
	Tissue Procurement: ATCC		
	CDNA Library Preparation: Life Technologies, Inc.		
	CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)		
	Clone sequencing by: Incyte Genomics, Inc.		
	CDNA distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.nih.gov		
	Plate: LHAM0429 row: m column: 13		
FEATURES	High quality sequence stop.		
SOURCE	Location/Qualifiers		
	1..915		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:4524852"		
	/clone_lib="NIH_MGC_92"		
	/tissue_type="embryonal carcinoma, cell line"		
	/lab_host="DH10B (phage-resistant)"		
	/note="Organ: testis; Vector: pCMV-SPO6; Site:1: NotI; Site:2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 2.5 kb. Library enriched for full-length clones and constructed by Life Technologies Note: this is a NIH_MGC library."		
BASE COUNT	246 a 252 c 227 g 190 t		
ORIGIN			
Alignment Scores:			
Pred. No.:	9 53e-89	Length:	915
Score:	975.50	Matches:	215
Percent Similarity:	77.36%	Conservative:	14
Best Local Similarity:	72.64%	Mismatches:	42
Query Match:	39.91%	Indels:	28
DB:	12	Gaps:	5

```

QY 1 MetalaSerThrIleuylSGluAlaLeuSerValValSerGluAspGlnSerLeuPheGlu 20
Db 59 ATGGCCAGCAGTATTAAAGGAAGCCCTTATCATCTTGTAGTGAGGACGAGCTGTTTGAG 118
QY 21 CysAlaIleuylSerProHisLeuAlaLysThrGluMetThrAlaSerSerSerGlu 40
Db 119 TGTCCTCAAGGACGACACACCTGCTGCTAGACAGAGATGACCGCTCTCTCCAGGCAC 178
QY 41 TyrlGlnIleuThrSerLysMetSerProArgValProGlnGlnAspTrpLeuSerGlnPro 60
Db 179 TATGGACAGAGCTTCCAGATAGAGCCCAAGCGCTCCCTCAGCAGATGGCTGTCCAAACC 238
QY 61 ProAlaArgValThrIleuLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsn 80
Db 239 CCAGCCAGGCGTCAACCAAAATGGAATGTACCTGACCGAGTGAATGGCTCAAGGAC 298
QY 81 SerProAspAspCysSerValAlaLysGlyGlyLysMetValSerSerSerAspAsnVal 100
Db 299 TCTCTGATGATGACGAGTGGCCAAAGCGGGAAGATGGGGGACACCCAGACACCGTT 358
QY 101 GlyMetAsnTyrlGlySerTyrlMetGluGluLysHisIleProProAsnMetThrThr 120
Db 359 GGGATGAACTACGCGACGCTACATGAGAGACACATGCCACC-CCAAACATCACCAGC 417
QY 121 AsnGluArgValIleValProAla-AspProThrLeuTrpSerThrAspHisValAr 140
Db 418 AACGAGGCGAGATATCGTGGCCAGCAAGATCTTACGCTATGAGTACAGACATGTGCG 477
QY 140 GGLnTrpLeuGlnTrpAlaValLysGluTyrlGlyLeuProAspAlaAspIleLeuLys 160
Db 478 GCACTGCTGAGTGGCGGCGTGAAGATATGGCTTCCAGACGCTCAACATCTTGTATT 537
QY 160 eGlnAsnIleAspGlyLysGluLeuLysMetThrLysAspAspPheGlnArgLeuTh 180
Db 538 CCAGAACATGATGAGGAAGGAAGCTGCAAGATGACCAAGACGACCTTCCAGAGGCTCAC 597
QY 180 rProSerTyrlAsnAlaAspIleLeuLysHisLeuHisIleuArgGluArg---G1 199
Db 598 CCCAGCTACAAATGCCACAT-CTTCTCTCAATCTCAGCTCAGAGAGACTCTCTG 656
QY 199 yAlaThrPheIlePhePro-----AsnThrSerValTyrlProGlnAlaThrGlnArgI 217
Db 657 TTCACATTTGACTTCAGATGATGTTGATTAACGCTTCAAAACCTCCAGCTTAATGC 716
QY 217 eThrThrArgProAspLeuProTyrlGlnAlaArgArgSerAlaTrpThrSerHisSe 237
Db 717 ATGCTAGAAACACAGATTATACATATGAGCCGCCAGAGATCAAGGCTGACGCTCACGG 776
QY 237 rHisProThrGlnSerLysAlaThrGlnProSerSerSerThrValProLysThrGlu-- 256
Db 777 -CACCCAAAGCCGACGTCAAAGCTGTCAATCTCTTACAGTGCCTCAATGAAGACAGTCT 835
QY 257 -----AspGlnArgProGlnLeuAspProTy 265
Db 836 CGGTAATCTATGATCTTGGCAGACAGTTCGCTGCAATGGGGCGGAGCTTGATGCC-- 893
QY 265 rGlnIleLeuGlyProThrSerSerArgLeuAlaAsnProGlySer 280
Db 894 -----GGGCGGGGGAGC-----CCGGGCTCC 914

```

RESULT 11
AUI36709 701 bp mRNA linear EST 02-AUG-2002
LOCUS AUI36709 PLACE1 Homo sapiens cDNA clone PLACE1004911 5', mRNA
DEFINITION sequence.
ACCESSION AUI36709
VERSION AUI36709.1 GI:10997248
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

```

REFERENCE 1 (bases 1 to 701)
AUTHORS Oca,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isogai,T.
TITLE HRI human cDNA project
JOURNAL Unpublished (2000)
COMMENT Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3975
Fax: 81-438-52-3986
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
FEATURES
Location/Qualifiers
source 1..701
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="PLACE1004911"
/tissue_type="placenta"
/note="Vector: pME18SF3"
BASE COUNT 199 a 185 c 166 g 145 t 6 others
ORIGIN
Alignment Scores:
Pred. No.: 1,62e-88 Length: 701
Score: 971.50 Matches: 185
Percent Similarity: 94.15% Conservative: 8
Best Local Similarity: 90.24% Mismatches: 10
Query Match: 39.75% Indels: 2
DB: 9 gaps: 1
US-09-902-772-2 (1-451) x AUI36709 (1-701)
QY 1 MetalaSerThrIleuylSGluAlaLeuSerValValSerGluAspGlnSerLeuPheGlu 20
Db 85 ATGGCCAGCAGTATTAAAGGAAGCCCTTATCATCTTGTAGTGAGGACGAGCTGTTTGAG 144
QY 21 CysAlaIleuylSerProHisLeuAlaLysThrGluMetThrAlaSerSerSerGlu 40
Db 145 TGTCCTCAAGGACGACACACCTGCTGCTAAGACAGATGACCGGCTCTCTCCAGGCAGC 204
QY 41 TyrlGlnIleuThrSerLysMetSerProArgValProGlnGlnAspTrpLeuSerGlnPro 60
Db 205 TATGGACAGAGCTTCCAGATAGAGCCCAAGCGCTCCCTCAGCAGATGGCTGTCTCAACC 264
QY 61 ProAlaArgValThrIleuLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsn 80
Db 265 CCAGCCAGGCGTCAACCAAAATGGAATGTACCTTACCGCGAGTGGCTCAAGGAC 324
QY 81 SerProAspAspCysSerValAlaLysGlyLysMetValSerSerSerAspAsnVal 100
Db 325 TCTCTGATGATGACGAGTGGCCAAAGCGGGAAGATGGGGGACGCCAGACACCGTT 384
QY 101 GlyMetAsnTyrlGlySerTyrlMetGluGluLysHisIleProProAsnMetThrThr 120
Db 385 GGGATGAACTACGCGACGCTACATGAGAGACAGACATGCCGCCAAACATGACACAGC 444
QY 121 AsnGluArgValIleValProAlaAspProThrLeuTrpSerThrAspHisValArg 140
Db 445 AACGAGGCGAGATATCGTGGCCAGAGATCTTACGCTATGAGATACAGACATGTGCGG 504
QY 141 GlnTrpLeuGlnTrpAlaValLysGluTyrlGlyLeuProAspValAspIleLeuLysPhe 160
Db 505 CAGTGGCTGAGTGGGGGGAAGATATGAGCTTCCAGACGTCACATCTGTATTTC 564
QY 161 GlnAsnIleAspGlyLysGluLeuLysMetThr-LysAspAspPheGlnArgLeuTh 180
Db 565 CAGAACATCGATGGAAGAACTGTGCAAGATGACCAAGAGACGATTCNANGACTCAA 624

```

Oy	180		ProterytRNAAlaAapIleLeuSerHisTyrLeuArgGlyAla	200
Dd	625	CCCCAGCTACAAACGCCGCACATCCTTCCTCACATCTCCACTACTC	---	ANGAGAGACTC 681
Oy	200	aThrPheIlePhe	204	
Dd	682	CTCTTTCACATTT	694	
RESULT	1.2			
LOCUS	AW948986		629 bp	mRNA linear EST 31-MAY-2000
DEFINITION	QV4-FR70005-110500-201-f02 FT0005 Homo sapiens cDNA, mRNA sequence.			
ACCESSION	AW948986			
VERSION	AW948986.1	GI:8126760		
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 629) Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.B., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J. Shotgun sequencing of the human transcriptome with ORF expressed sequence tags Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)			
JOURNAL	MEDLINE			
COMMENT	Contact: Stimpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel.: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?lt=st2-QV4-FR70005-110 500-201-f02&t3=2000-05-11&t4=1) Seq primer: puc 18 forward High quality sequence stop: 628. Location/Qualifiers 1..629 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_lib="FR70005" /dev_stage="Adult" /note="Organ: prostate,tumor; Vector: puc18; Site:1; SmaI: Site-2; SmaI: A mini-library was made by cloning products derived from ORESTS PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."			
FEATURES	source			
BASE COUNT	166 a	173 c	171 g	119 t
ORIGIN				
Alignment Scores:	Pred. NO.:	Length:	629	
	Score:	Matches:	180	
	Percent Similarity:	Conservative:	7	
	Best local Similarity:	Mismatches:	5	
	Query Match:	Indels:	0	
	DG:	Gaps:	0	
US-09-902-772-2 (1-451) x AW948986 (1-629)				
Oy	6	LysGluAlaLeuSerValValSerLysAspGlnSerLeuPheGluCysAlaTyrGlySer	25	
	:	::::	:	

D	b	19	CAGGAAGCCCTTATCAGTTGTGATGAGGACCAAGTCGTTGTATTGAGTGCTACGGAACG	78
Oy		26	ProHslsleuAlaLysThrGlumethrAlaSerSerSerSelutyrGlylnhrser	45
D	b	79	CCACACTGGCGCTAAACAAGAGATGACC GGCTCTCTCTCCAGCATATGACAGACTTCC	138
Oy		46	LysMetSerProArValProGlnInlAspTrPLeuSerGlnProProAlaArgValThr	65
D	b	139	AAGATGAGCCCCACGGCTCCCTCAGCAGATTTGGCTGTCTCAACCCCCCAGGGTCAAC	198
Oy		66	IleLysMetGluCysasnProAsnGlnValAsnGlySerArgAsnSerProAspAspCys	85
D	b	199	ATCAAATGAATGTAAACCTTAGCCAGGAGTGAATGGCTCAAGAACTCTCCTGATGAATGC	258
Oy		86	SerValAlaLysGlyGlyLysMetValSerSerSerAspAsnValGlyMetAsnTyrGly	105
D	b	259	AGTGGGCCAAGGGGGAGAAGATGGTGGGACACCAGAACCCGTTGGAGTAACATCACGC	318
Oy		106	SerTyrMetGlnGlnLysHisIleProProProAsnMetThrThrAsnGluArgArgVal	125
D	b	319	AGCTACATGAGAGAGACACATGCCACCCCACAAAACATGACCAGCAGAGGCGAGAGTT	378
Oy		126	IleValProAlaAspProThrLeuTyrSerThrAspHisValArgGlnTyrLeuGluTyr	145
D	b	379	ATCGTGCCAGCAGAGATCCTACGCTATGAGATACAGACATGTGGCGACATGGCTGAGTGG	438
Oy		146	AlaValLysGluTyrGlyLeuProAspValAspIleLeuLeuPheGlnAsnIleaspGly	165
D	b	439	GCGGTGAAGAATATATAGCGCTTCCACACCGCTCAACTCTTGTATTCACAAACATCGATGG	498
Oy		166	LysGluLeuCysAllyMetThrLysAspAspPheGlnArgLeuThrProSerTyrAsnAla	185
D	b	489	AAGGACAGCTGGCAAGATACCAATACACATCTCCAGAGGCTCACCCAGCATCAACGCC	558
Oy		186	AspIleLeuLeuSerHisLeuHisTyrLeuArgGln	197
D	b	559	GACATCCTTCTCTCATCTCCATCTCCATCACTCAGAGAG	594
RESULT_13				
BQ953977			1074 bp	mRNA linear EST 21-AUG-2002
LOCUS				
DEFINITION			AGENCODEUTR.8877989 Lupskl.sciatric.nerve Homo sapiens cDNA clone	
ACCESSION			BQ953977	
VERSION			BQ953977.1	GI:22369455
KEYWORDS			EST.	
SOURCE			human.	
ORGANISM			Homo sapiens	
REFERENCE			Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS			Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.	
TITLE			1 (bases 1 to 1074)	
JOURNAL			NHR-MGC http://mgc.nci.nih.gov/.	
COMMENT			National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgapbs-re@mail.nih.gov Tissue Procurement: Dr. James R. Lupski CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LMNL at: http://image.llnl.gov Plate: LMAM13610 row: h column: 17 High quality sequence stop: 575. Location/Qualifiers 1..1074 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:6198592" /clone_lib="Lupskl.sciatric.nerve" /sex="male" /tissue_type="sciatric nerve"	
FEATURES				
Source				

us-09-902-772-2 (1-451) x BM489636 (1-634)

```
OY 120 ThrAsnGluArgValIleValProAlaAspProThrLeuTrpSerThrAspHisVal 139
|||||
Db 3 ACCAATGAACGAAGAGTATTGTGCCACAGATCTACGTTATGGAGCAGACCATGTA 62
|||||
OY 140 ArgGlnTrpLeuGlnTrpAlaValLysGluTyrGlyLeuProAspValAspIleLeu 159
|||||
Db 63 CGGCGTGGCTGGAGTGGGCAAGAGTATGGTCTCCAGACGTGNCATCTGTGTTG 122
|||||
OY 160 PheGlnAsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArgLeu 179
|||||
Db 123 TTCCGAACATTGATGAGGAAGAGTGTGTAAATGACCAAGATGACTCCAGAGACTC 182
|||||
OY 180 ThrProSerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGluArg--- 198
|||||
Db 183 ACGCCGAGCTATACGACAGATATCTCTGTCACACACTACACTCAGAGAGACTCCT 242
|||||
OY 198 ----- 198
Db 243 CTTCACATTGACTTCAGATGATGTGATAGGCCCTTACAAACTCTCCAGGTTAATG 302
OY 199 -----GlyAlaThrPheIlePheProAsnThrSerValTyrProGlu 212
Db 303 CATGCTAGAAACACAGAGAGGCCACTTTTATTTTCCAAATACATCAGTTTACCCAGAA 362
OY 213 AlaThrGlnArgIleThrThrArgProAspLeuProTyrGluGlnAlaArgSerAla 232
|||||
Db 363 GCAACGCAAGAAATPAAACAAGAGCCAGATTACTTATGAGCAAGCAGAGGATCAGCG 422
|||||
OY 233 TrpThrSerHisSerHisProThrGlnSerLysAlaThrGlnProSerSerThrVal 252
|||||
Db 423 TGGACGAGTACAGCCATCCATCAGTCAAAAGCTACCAACCATCTTCAACAGGTG 482
|||||
OY 253 ProLysThrGluAspGlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSer 272
|||||
Db 483 CCCAAACAGAAAGACACAGCGTCTCTAGTTAGATCCTTATCAGATTCTTGACCGACAGC 542
|||||
OY 273 SerArgLeuAlaAsnProGlySerGlySerGlnIleGlnLeuTrpGlnPheLeuLeuGluLeu 292
|||||
Db 543 AGCGGTCTTGCAAAATCCAGGAGGTGGCAGATACAGCTATGCAAGTCTCTACTGAGGCTT 602
|||||
OY 293 LeuSerAspSerSerAsnSerAsnCysIle 302
|||||
Db 603 CTGTGAGACAGTCCCACTCACTGCAATC 632
|||||
```

Search completed: November 9, 2002, 19:46:44
Job time : 1512.44 secs

THIS PAGE BLANK (USPTO)

PT deformans or ossification of spinal column ligament
XX
PS Claim 5; Page 9-10; 15pp; Japanese.
XX
CC The present sequence represents chicken c-ery protein. The specification
CC also describes a chicken C-11 protein (AA01520) which has cell
CC calcification inhibiting activity and a cell calcification inhibiting
CC agent containing c-ery protein. The proteins are used for measuring the
CC calcification of a cell, for diagnosing arthritis deformans or
CC ossification of spinal column ligament.
XX
SQ Sequence 478 AA;
SO
Query Match 100.0%; Score 2588; DB 20; Length 478;
Best Local Similarity 100.0%; Pred. No. 1,4e-227;
Matches 478; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MASTIKALSVSEDSLEFCAYGSPHLAKTEMTASSSEYQTSKMSRVPQOQMLSP 60
DB 1 MASTIKALSVSEDSLEFCAYGSPHLAKTEMTASSSEYQTSKMSRVPQOQMLSP 60
QY 61 PARVITKMECNPNQVNGSNRSPDCSVAGKGVSSSDVNGMNGSYMEKHIPEPNTMT 120
DB 61 PARVITKMECNPNQVNGSNRSPDCSVAGKGVSSSDVNGMNGSYMEKHIPEPNTMT 120
QY 121 NRRRYIVPADPILMSIDHVRQWLEMAVKEYGLPDVDILFQNIIDKEICKMKRDFQRLT 180
DB 121 NRRRYIVPADPILMSIDHVRQWLEMAVKEYGLPDVDILFQNIIDKEICKMKRDFQRLT 180
QY 181 PSYNDILSHLHYLEREPLPHLITSDVDYKALONSPRLMHAHNTGGATPIFPNTSVYPPA 240
DB 181 PSYNDILSHLHYLEREPLPHLITSDVDYKALONSPRLMHAHNTGGATPIFPNTSVYPPA 240
QY 241 TORITTRPDLPEQARRSAMTSHPTQSKATOPSSSYVKTEDORPOLDPYQIIGPTSS 300
DB 241 TORITTRPDLPEQARRSAMTSHPTQSKATOPSSSYVKTEDORPOLDPYQIIGPTSS 300
QY 301 RIANGSGGQIOLOMFLLELSDSSNSNCITWGTNGEEMKTPDEYARRWGERKSKPNN 360
DB 301 RIANGSGGQIOLOMFLLELSDSSNSNCITWGTNGEEMKTPDEYARRWGERKSKPNN 360
QY 361 YKLSRALRYYYDKNIMTVHPPESSMYKPPSDLPYMSYHGKRYAKKDFPGIAOALDP 420
DB 361 YKLSRALRYYYDKNIMTVHPPESSMYKPPSDLPYMSYHGKRYAKKDFPGIAOALDP 420
QY 421 HAHPOKMFVADHPALPVYSSFFAAPRPYNSPTGGIYPTRLPAHMPSHLGTYY 478
DB 421 HAHPOKMFVADHPALPVYSSFFAAPRPYNSPTGGIYPTRLPAHMPSHLGTYY 478
RESULT 2
AA01520
ID AA01520 standard; Protein; 451 AA.
AC AA01520;
XX
DI 23-JUN-1999 (first entry)
XX
DE Chicken C-11 protein.
XX
KW Chicken; C-11 protein; cell calcification inhibiting activity;
KW cell calcification inhibiting agent; c-ery protein; arthritis deformans;
KW ossification; spinal column ligament.
XX
OS Gallus sp.
XX
PN JP11075871-A.
XX
PD 23-MAR-1999.
XX
PF 29-MAY-1998; 98JP-0166076.
XX
PR 20-JUN-1997; 97US-0050297.

PR 18-JUN-1997; 97US-0878177.
XX
XX (CHUS) CHUGAI PHARM CO LTD.
PA (UYPE-) UNIV PENNSYLVANIA.
XX
XX WPI: 1999-257708/22.
DR N-PSDB; AAX26551.
XX
XX An active protein for inhibiting cell calcification - useful for
PT measuring the calcification of a cell, for diagnosing arthritis
PT deformans or ossification of spinal column ligament
XX
PS Claim 1; Page 8; 15pp; Japanese.
XX
CC The present sequence represents a chicken C-11 protein which has cell
CC calcification inhibiting activity. The specification also describes
CC a cell calcification inhibiting agent containing c-ery protein
CC (AA01521). The proteins are used for measuring the calcification of a
CC cell, for diagnosing arthritis deformans or ossification of spinal
CC column ligament.
XX
SQ Sequence 451 AA;
SO
Query Match 87.4%; Score 2262.5; DB 20; Length 451;
Best Local Similarity 86.3%; Pred. No. 7,4e-198;
Matches 430; Conservative 0; Mismatches 1; Indels 67; Gaps 3;
QY 1 MASTIKALSVSEDSLEFCAYGSPHLAKTEMTASSSEYQTSKMSRVPQOQMLSP 60
DB 1 MASTIKALSVSEDSLEFCAYGSPHLAKTEMTASSSEYQTSKMSRVPQOQMLSP 60
QY 61 PARVITKMECNPNQVNGSNRSPDCSVAGKGVSSSDVNGMNGSYMEKHIPEPNTMT 120
DB 61 PARVITKMECNPNQVNGSNRSPDCSVAGKGVSSSDVNGMNGSYMEKHIPEPNTMT 120
QY 121 NRRRYIVPADPILMSIDHVRQWLEMAVKEYGLPDVDILFQNIIDKEICKMKRDFQRLT 180
DB 121 NRRRYIVPADPILMSIDHVRQWLEMAVKEYGLPDVDILFQNIIDKEICKMKRDFQRLT 180
QY 181 PSYNDILSHLHYLEREPLPHLITSDVDYKALONSPRLMHAHNTGGATPIFPNTSVYPPA 240
DB 181 PSYNDILSHLHYLEREPLPHLITSDVDYKALONSPRLMHAHNTGGATPIFPNTSVYPPA 240
QY 241 TORITTRPDLPEQARRSAMTSHPTQSKATOPSSSYVKTEDORPOLDPYQIIGPTSS 300
DB 241 TORITTRPDLPEQARRSAMTSHPTQSKATOPSSSYVKTEDORPOLDPYQIIGPTSS 300
QY 301 RIANGSGGQIOLOMFLLELSDSSNSNCITWGTNGEEMKTPDEYARRWGERKSKPNN 360
DB 301 RIANGSGGQIOLOMFLLELSDSSNSNCITWGTNGEEMKTPDEYARRWGERKSKPNN 360
QY 361 YKLSRALRYYYDKNIMTVHPPESSMYKPPSDLPYMSYHGKRYAKKDFPGIAOALDP 420
DB 361 YKLSRALRYYYDKNIMTVHPPESSMYKPPSDLPYMSYHGKRYAKKDFPGIAOALDP 420
QY 401 HGRKRYAKKDFPGIAOALDPAAHPOKMFVADHPALPVYSSFFAAPRPYNSPTGGIY 460
DB 401 HGRKRYAKKDFPGIAOALDPAAHPOKMFVADHPALPVYSSFFAAPRPYNSPTGGIY 460
QY 461 PNTRLPAHMPSHLGTYY 478
DB 461 PNTRLPAHMPSHLGTYY 478
RESULT 3
AAU75313
ID AAU75313 standard; Protein; 462 AA.
AC AAU75313;
XX
DI 23-APR-2002 (first entry)
XX
DE Human Ery transcription factor.

Human; Ery; transcription factor; tumour necrosis factor- α ; TNF- α ; interleukin-1 (IL-1); SPARC; thrombospondin; TSP; rho-A; intracellular adhesion molecule-2; ICAM-2; von Willebrand Factor; vWF; angiolipid; inflammatory disorder; developmental disorder; wound; menstrual disorder; cancer; rheumatoid arthritis; atherosclerosis; diabetic retinopathy; stenosis; osteoporosis; cataract; diabetes mellitus; glomerulonephritis; inflammatory glomerular disease; vasculitis; retinopathy; liver fibrosis; haemoglobinopathy; Crohn's disease.

Homo sapiens.

WO20018122-A2.

22-NOV-2001.

15-MAY-2001; 2001WO-GB02140.

16-MAY-2000; 2000US-204493P.

(GLAXO) GLAXO GROUP LTD.

McCaughlin F, Randl AM; WPI; 2002-089855/12.

New modulator of a member of the Ets family of transcription factors, Ery, which is not tumour necrosis factor- α or interleukin-1, for treating cancer, rheumatoid arthritis, atherosclerosis, stenosis, and osteoporosis -

Disclosure; Page 53-54; 75pp; English.

The invention relates to an Ery (a member of the Ets family of transcription factors) modulator for use in medicine, where the modulator is not tumour necrosis factor- α (TNF- α) or interleukin-1 (IL-1). Also included are a method of screening for a drug or drug candidate, involving determining whether or not a group being screened modulates the effect of Ery on secreted protein, acidic, cysteine-rich (osteocalcin/BM-40) (SPARC), thrombospondin (TSP), rho-A (a Ras superfamily member), intracellular adhesion molecule-2 (ICAM-2) and/or von Willebrand Factor (vWF) RNA or polypeptide expression, or expression of heterologous RNA or a heterologous polypeptide that is under the control of one or more regulatory sequences of the SPARC, TSP, rho A, ICAM-2 and/or vWF genes, determining whether or not a group being screened modulates the expression of Ery or of Ery RNA in a given expression system, determining whether or not a group being screened modulates binding of Ery to a nucleic acid comprising an Ery binding site, determining whether or not a nucleic acid hybridises with DNA of an Ery gene or its transcription product or determining whether or not a nucleic acid knocks out or reduces expression of Ery. Ery based probes can be used in the diagnosis of a disorder such as a disorder in which angiogenesis is involved, inflammatory disorder, developmental disorder, wound, menstrual disorder involving cell proliferation, a disorder that can be treated by tissue remodeling, cancer, cancer metastasis, rheumatoid arthritis, atherosclerosis, diabetic retinopathy, stenosis, osteoporosis, cataract, diabetes mellitus, glomerulonephritis, inflammatory glomerular disease, vasculitis, retinopathy, liver fibrosis, haemoglobinopathy or Crohn's disease. The modulator is useful in the preparation of a medicament for treating a disorder which involves aberrant expression of SPARC, TSP, rho A, ICAM-2 or vWF. A group that modulates the effect of Ery or a therapeutically active agent is useful for the preparation of a medicament for treating the above mentioned disorders. The present sequence is the amino acid sequence of human Ery.

Oy	2	ASTIKELSVYSEDQSLFECAAGSPHLATTEMTASSSEVOTSKMSRPVPODDMLSQP
Oy	9	AAHKELSVYSEDQSLFECAAGTPHLATTEMTASSSSSDYGOTSKMSRPVPODDMLSQP
Oy	62	ARVITKMECPNPOVNGSRNSPDDCSYAKGGKAVSSSDVNGVNGYSYAEKHIPPNNMTT
Db	69	ARVITKMECPNPOVNGSRNSPDDCSYAKGGKAVGSPDTVGNNGYSYAEKHMPNNMTT
Oy	122	ERRIYVADPTLWSTDHYNQMLEMAVKEGLEDVDVILLFQNDKEKCKTKTDFORLTF
Db	129	ERRIYVADPTLWSTDHYNQMLEMAVKEGLEDVDVILLFQNDKEKCKTKTDFORLTF
Oy	182	SYNADILSLHLHLRETPLEPHLITSDVDVDAALQNSPRLMHANNTGATFIPNTSVPEAT
Db	189	SYNADILSLHLHLRETPLEPHLITSDVDVDAALQNSPRLMHANNT-----
Oy	242	QRITRPDLPEQARNSAMTSSHSPT-OSKATOPSSSTVPKTEQDRPOLDPYQILGPTSS
Db	232	-----DLPEEPERRSASMTAGCHGHPPOSKAOPSPSTVPKTEQDRPOLDPYQILGPTSS
Oy	301	RLAPGSGQIQLOMQLFLELISDSSNCTTGTGTEGEEKMTDPEVARRGERSKPNMNA
Db	285	RLAPGSGQIQLOMQLFLELISDSSNCTTGTGTEGEEKMTDPEVARRGERSKPNMNA
Oy	361	YDKLSRALRYYYKKNIMTKVHPRESSMYKYPEDLPYMSYHGKKRYAAKFPDHGIALQAP
Db	345	YDKLSRALRYYYKKNIMTKV-----HGKKRYAAKFPDHGIALQAP
Oy	421	-----HAHPQKMFVAPHPALPYTSSSFFAAPNPYMSPTGIGTY
Db	385	HPRESSLYKKYSDLPYMGSYHAHPQKMFVAPHPALPYTSSSFFAAPNPYMSPTGIGTY
Oy	461	PNTRLPAAHMPSHLGTYT 478
Db	445	PNTRLPTSHMPSHLGTYT 462
RESULT 4		
AAR44556		
ID	AAR44556	standard; Protein; 452 AA.
XX	AAR44556;	
AC		
XX		
DT	26-MAY-1994	(first entry)
XX		
DE		
XX		
KW	Human HMV-FLI-1 gene product.	
KW	chromosomal translocation; chimeric; chimaeic; Ewing sarcoma;	
KW	Ews gene; malignant melanoma; hum-flt-1; clone BW025;	
KM	primitive peripheral neuroectodermal tumour; human chromosome 11;	
XX	human chromosome 22; ss.	
XX		
OS	Homo sapiens.	
NN	MO9323549-A.	
XX		
PD	25-NOV-1993.	
XX		
PF	19-MAY-1993;	93WO-FR00494.
XX		
PR	20-MAY-1992;	92FR-0006123.
XX		
XX		
FA	(CNRS) CNRS CENT NAT RECH SCT.	
PI	Aurias A, Delattre O, Desmaza C, Melot T, Peter M;	
PI	Plougastel B, Thomas G, Zucman J;	
XX		
DR	WPI: 1993-386580/48.	
XX	N-PSDB: AAQ50644.	
XX		
PT	New nucleic acid of EWS gene and its hybrid(s) - contg. gene	
PT	sequence involved in chromosomal trans-location, also derived	
PT	mRNA, probes, fusion proteins etc., for diagnosis and treatment	

XX Peptide #1701 encoded by breast cell single exon nucleic acid probe.
DE Human; microarray; single exon probe; gene expression; breast;
KW disease; cancer.
XX Homo sapiens.
OS
XX MO200157271-A2.
PN
XX 09-AUG-2001.
PD
XX 30-JAN-2001; 2001WO-US00662.
PF
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI
XX WPI; 2001-496933/54.
DR
XX
XX New spatially-addressable set of single exon nucleic acid probes,
PT useful for measuring gene expression in sample derived from human
PT breast, comprises number of single exon nucleic acid probes -
XX
XX Claim 27; SEQ ID NO 12018; 327pp + sequence listing; English.
PS
XX
XX The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human breast and BT 474 cells. The method involves contacting
CC the probes with a collection of detectably labelled nucleic acids
CC derived from mRNA of human breast, and then measuring the label
CC bound to each probe of the microarray. The probes are useful for
CC verifying the expression of regions of genomic DNA predicted to
CC encode proteins. They are useful for gene discovery and for
CC determining predisposition and/or prognosing breast disease. Gene
CC expression analysis is useful for assessing the toxicity of chemical
CC agents on cells. The microarray of this invention presents a far greater
CC diversity of probes for measuring gene expression, with far less bias
CC than expressed sequence tag microarrays. The method is suitable for
CC rapid production of functional information from genomic sequence. The
CC present sequence is a peptide encoded by a single exon nucleic acid
CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 188 AA:
SQ
Query Match 31.0%; Score 801; DB 22; Length 188;
Best Local Similarity 76.6%; Pred. No. 7.9e-65;
Matches 151; Conservative 3; Mismatches 3; Indels 40; Gaps 2;
QY 302 LANCGSGIOIQMOPFLELLSDSSNSNCITWGTNGEFKMTDPDEYARRWGERKSKPPNNMY 361
DB 12 LTLSSGSGIOIQMOPFLELLSDSSNSNCITWGTNGEFKMTDPDEYARRWGERKSKPPNNMY 71
QY 362 DKLSRALRYYYDKNIMTVHPPESSMYKYPDDLPMSSYHGKRYAYKRPFGIAQALQP- 420
DB 72 DKLSRALRYYYDKNIMTVHPPESSMYKYPDDLPMSSYHGKRYAYKRPFGIAQALQP 111
QY 421 -----HAHPQKMFVAPHPALPYTSSSFFAAPNPYNNSPGTG1YP 461
DB 112 PPESSLYKYPDDLPMSSYHAPQKMFVAPHPALPYTSSSFFAAPNPYNNSPGTG1YP 171
QY 462 NTRLPAAHMPSHLGTYY 478

DB 172 NTRLPAAHMPSHLGTYY 188
|||||:|||||
RESULT 7
ABB34212
ID ABB34212 standard; Peptide: 188 AA.
XX
XX ABB34212;
AC
XX 04-FEB-2002 (first entry)
DE
XX Peptide #1718 encoded by human foetal liver single exon probe.
XX Human; foetal liver; gene expression; single exon nucleic acid probe.
XX
XX Homo sapiens.
OS
XX MO200157277-A2.
PN
XX 09-AUG-2001.
PD
XX 30-JAN-2001; 2001WO-US00669.
PF
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI
XX WPI; 2001-483447/52.
DR
XX
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human foetal liver -
XX
XX Claim 27; SEQ ID NO 26847; 639pp + sequence listing; English.
PS
XX
XX The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC foetal liver. The present sequence is a peptide encoded by a single exon
CC nucleic acid probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 188 AA:
SQ
Query Match 31.0%; Score 801; DB 22; Length 188;
Best Local Similarity 76.6%; Pred. No. 7.9e-65;
Matches 151; Conservative 3; Mismatches 3; Indels 40; Gaps 2;
QY 302 LANCGSGIOIQMOPFLELLSDSSNSNCITWGTNGEFKMTDPDEYARRWGERKSKPPNNMY 361
DB 12 LTLSSGSGIOIQMOPFLELLSDSSNSNCITWGTNGEFKMTDPDEYARRWGERKSKPPNNMY 71
QY 362 DKLSRALRYYYDKNIMTVHPPESSMYKYPDDLPMSSYHGKRYAYKRPFGIAQALQP- 420
DB 72 DKLSRALRYYYDKNIMTVHPPESSMYKYPDDLPMSSYHGKRYAYKRPFGIAQALQP 111
QY 421 -----HAHPQKMFVAPHPALPYTSSSFFAAPNPYNNSPGTG1YP 461
DB 112 PPESSLYKYPDDLPMSSYHAPQKMFVAPHPALPYTSSSFFAAPNPYNNSPGTG1YP 171
QY 462 NTRLPAAHMPSHLGTYY 478

DB 172 NTRLPESHMPSHLGTTY 188

RESULT 8
ABBI9649 standard; Protein; 188 AA.

ID ABB19649 standard; Protein; 188 AA.

XX ABB19649;

AC ABB19649;

XX 23-JAN-2002 (first entry)

DT 23-JAN-2002 (first entry)

XX Protein #1648 encoded by probe for measuring heart cell gene expression.

DE Human; gene expression; heart; microarray; vascular system;

XX Human; gene expression; heart; microarray; vascular system;

KW cardiovascular disease; hypertension; cardiac arrhythmia;

KW congenital heart disease.

XX

OS Homo sapiens.

XX

PN WO200157274-A2.

XX 09-AUG-2001.

PD 09-AUG-2001.

XX

XX 30-JAN-2001; 2001WO-US00666.

PF 30-JAN-2001; 2001WO-US00666.

XX

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX

PA (MOLE-) MOLECULAR DYNAMICS INC.

XX

PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488899/53.

DR WPI; 2001-488899/53.

XX

PT Single exon nucleic acid probes for analyzing gene expression in human

PT hearts -

XX

PS Claim 15; SEQ ID No 21419; 530pp; English.

XX

XX The present invention relates to single exon nucleic acid probes for

CC measuring human gene expression in a sample derived from human heart (see

CC ABA21535-ABA41305). The present sequence is a protein encoded by one such

CC probe. The probes may be used for predicting, measuring and displaying

CC gene expression in samples derived from the human heart via microarrays.

CC By measuring gene expression, the probes are useful for predicting,

CC diagnosing, grading, staging, monitoring and prognosing diseases of the

CC human heart and vascular system e.g. cardiovascular disease,

CC hypertension, cardiac arrhythmias and congenital heart disease.

CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pcl_sequences.

XX

XX Sequence 188 AA;

Query Match 31.0%; Score 801; DB 22; Length 188;

Best Local Similarity 76.6%; Pred. No. 7.9e-65;

Matches 151; Conservative 3; Mismatches 3; Indels 40; Gaps 2;

QY 302 LANPGSGOIQIOMQFLLELLSDSSNSNCITWEGTNGEFKMTDPDEVARRMGERKSKPMNY 361

DB 12 LTLGSGOIQIOMQFLLELLSDSSNSNCITWEGTNGEFKMTDPDEVARRMGERKSKPMNY 71

QY 362 DKLSRALRYTDDKNIMTKYHPPESSMKTIYPSDLPYMSYHGKRYAYKFDHGTGAOLQP- 420

DB 72 DKLSRALRYTDDKNIMTKYHPPESSMKTIYPSDLPYMSYHGKRYAYKFDHGTGAOLQP 111

QY 421 -----HAHPQKMFVAPHPALPYTSSSFFAAPNPYWNSTPTGTIY 461

DB 112 PPSSLYKTYPSDLPYMGSTYAHHPQKMFVAPHPALPYTSSSFFAAPNPYWNSTPTGTIY 171

QY 462 NTRLPAAHMPSHLGTYY 478

DB 172 NTRLPESHMPSHLGTTY 188

RESULT 9
AAM55005 standard; Protein; 188 AA.

ID AAM55005 standard; Protein; 188 AA.

XX AAM55005;

AC AAM55005;

XX

DT 05-NOV-2001 (first entry)

DT 05-NOV-2001 (first entry)

XX Human brain expressed single exon probe encoded protein SEQ ID NO: 27110.

DE Human brain expressed single exon probe encoded protein SEQ ID NO: 27110.

XX

KW Human; brain expressed exon; gene expression analysis; probe;

KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;

KW epilepsy; cancer.

XX

OS Homo sapiens.

XX

PN WO200157275-A2.

XX 09-AUG-2001.

PD 09-AUG-2001.

XX

XX 30-JAN-2001; 2001WO-US00667.

PF 30-JAN-2001; 2001WO-US00667.

XX

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX

PA (MOLE-) MOLECULAR DYNAMICS INC.

XX

PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483446/52.

DR WPI; 2001-483446/52.

XX

PT Single exon nucleic acid probes for analyzing gene expression in human

PT brains -

XX

PS Example 4; SEQ ID NO: 27110; 650pp + Sequence listing; English.

XX

XX The present invention provides a number of single exon nucleic acid

CC probes which are derived from genomic sequences expressed in the human

CC brain. They can be used to measure gene expression in brain cell samples,

CC which may enable the diagnosis and improved treatment of nervous system

CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,

CC epilepsy and cancers. The present sequence is a protein encoded by one of

CC the probes of the invention.

XX

XX Sequence 188 AA;

Query Match 31.0%; Score 801; DB 22; Length 188;

Best Local Similarity 76.6%; Pred. No. 7.9e-65;

Matches 151; Conservative 3; Mismatches 3; Indels 40; Gaps 2;

QY 302 LANPGSGOIQIOMQFLLELLSDSSNSNCITWEGTNGEFKMTDPDEVARRMGERKSKPMNY 361

DB 12 LTLGSGOIQIOMQFLLELLSDSSNSNCITWEGTNGEFKMTDPDEVARRMGERKSKPMNY 71

QY 362 DKLSRALRYTDDKNIMTKYHPPESSMKTIYPSDLPYMSYHGKRYAYKFDHGTGAOLQP- 420

DB 72 DKLSRALRYTDDKNIMTKYHPPESSMKTIYPSDLPYMSYHGKRYAYKFDHGTGAOLQP 111

QY 421 -----HAHPQKMFVAPHPALPYTSSSFFAAPNPYWNSTPTGTIY 461

DB 112 PPSSLYKTYPSDLPYMGSTYAHHPQKMFVAPHPALPYTSSSFFAAPNPYWNSTPTGTIY 171

OY 462 NTRLPAHMPSHLGTYY 478
 DB 172 NTRLPTSHMPSHLGTYY 188

RESULT 12

ID AAM27682
 AAM27682 standard; Protein; 188 AA.

AC AAM27682;

DT 17-OCT-2001 (first entry)

DE Peptide #1719 encoded by probe for measuring placental gene expression.

KW Probe; microarray; human; placenta; antenatal diagnosis;
 genetic disorder.

OS Homo sapiens.

PN WO200157272-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00663.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

PA (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI; 2001-488897/53.

PT Human genome-derived single exon nucleic acid probes useful for

PT analyzing gene expression in human placenta -

PS Claim 27; SEQ ID No 27951; 654pp; English.

XX The present invention relates to single exon nucleic acid probes (SENP;

XX see A113135-A1157546). The present sequence is a peptide encoded by one

XX CC such probe. The probes are useful for producing a microarray for

XX CC predicting, measuring and displaying gene expression in samples derived

XX CC from human placenta. The probes are useful for antenatal diagnosis of

XX CC human genetic disorders.

XX Sequence 188 AA;

Query Match 31.0%; Score 801; DB 22: Length 188;

Best Local Similarity 76.6%; Pred. No. 7.9e-65;

Matches 151; Conservative 3; Mismatches 3; Indels 40; Gaps 2;

RESULT 13

ID AAM02964
 AAM02964 standard; Protein; 188 AA.

AC AAM02964;

DT 09-OCT-2001 (first entry)

DE Peptide #1646 encoded by probe for measuring breast gene expression.

KW Probe; human; breast disease; breast cancer; development disorder;

KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.

OS Homo sapiens.

PN WO200157270-A2.

PD 29-JAN-2001; 2001WO-US00661.

PF 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

PA (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI; 2001-476286/51.

PT Novel single exon nucleic acid probe used to measuring gene expression

PT in a human breast -

PS Claim 27; SEQ ID No 11704; 322pp; English.

XX The present invention relates to novel single exon nucleic acid probes

XX (see A1100010-A1110067). The present sequence is a peptide encoded by one

XX CC such probe. The probes are useful for measuring human gene expression in

XX CC a human breast sample, where the probe hybridizes at high stringency to a

XX CC nucleic acid expressed in the human breast. The probes are useful for

XX CC predicting, diagnosing, grading, staging, monitoring and prognosing

XX CC diseases of the human breast, particularly those diseases with polygenic

XX CC aetiology. The diseases include: breast cancer, disorders of development,

XX CC inflammatory diseases of the breast, fibrocystic changes, proliferative

XX CC breast disease and non-carcinoma tumours.

XX CC Note: The sequence data for this patent did not form part of the printed

XX CC specification, but was obtained in electronic format directly from WIPO

XX CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 188 AA;

Query Match 31.0%; Score 801; DB 22: Length 188;

Best Local Similarity 76.6%; Pred. No. 7.9e-65;

Matches 151; Conservative 3; Mismatches 3; Indels 40; Gaps 2;

OY 302 LANPGSGQIQLOMQLLELSDSSNSNCITWEGTNGEFKMTDDEVARRRGERSKPNMNY 361

DB 12 LFSLSGSGQIQLOMQLLELSDSSNSNCITWEGTNGEFKMTDDEVARRRGERSKPNMNY 71

OY 362 DKLSRALRYTYDKNIMTKYHPPSSMYKYPSPDLYWSSYHGKRYAKKEDFHGIAQALQP- 420

DB 72 DKLSRALRYTYDKNIMTKY- HGRKRYAKKEDFHGIAQALQP 111

OY 421 HAHPOKMNFAVAPHPALPYTSSSFFAAPNPNYNSPTGCIYP 461

DB 112 PPESSLYKYPSPDLYWSSYHAPPOKMNFAVAPHPALPYTSSSFFAAPNPNYNSPTGCIYP 171

QY 462 NTRLPAAHPSHIGTTY 478
 DB 172 NTRLPAAHPSHIGTTY 188

RESULT 14
 ABG37017
 ID ABG37017 standard; Peptide; 188 AA.
 AC ABG37017;
 DT 19-AUG-2002 (first entry)
 DE Human peptide encoded by genome-derived single exon probe SEQ ID 26682.
 XX
 KW Human: single exon probe; asthma; lung cancer; COPD; ILD;
 KW chronic obstructive pulmonary disease; interstitial lung disease;
 KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
 KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
 KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
 KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;
 KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
 KW primary ciliary dyskinesia; pulmonary hypertension;
 KW hyaline membrane disease.
 XX
 OS Homo sapiens.
 PN WO200186003-A2.
 XX
 PD 15-NOV-2001.
 XX
 PE 30-JAN-2001; 2001WO-US00665.
 XX
 PR 04-FEB-2000; 2000US-180312P.
 PR 26-MAY-2000; 2000US-207456P.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-234687P.
 PR 27-SEP-2000; 2000US-236359P.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2002-114183/15.
 XX
 PT Spatially-addressable set of single exon nucleic acid probes, used to
 PT measure gene expression in human lung samples -
 XX
 PS Claim 27; SEQ ID NO 26682; 634P; English.
 XX
 CC The invention relates to a spatially-addressable set of single exon
 CC nucleic acid probes for measuring gene expression in a sample derived
 CC from human lung comprising single exon nucleic acid probes having one of
 CC 12614 nucleic acid sequences mentioned in the specification, or their
 CC complements or the 12387 open reading frames derived from the 12614
 CC probes. Also included are a microarray comprising the novel set of
 CC probes; the novel set of probes which hybridise at high stringency to a
 CC nucleic acid expressed in the human lung; measuring gene expression in a
 CC sample derived from human lung, comprising (a) contacting the array with
 CC a collection of detectably labeled nucleic acids derived from human lung
 CC mRNA, and (b) measuring the label detectably bound to each probe of
 CC the array; identifying exons in a eukaryotic genome, comprising
 CC (a) algorithmically predicting at least one exon from genomic sequences
 CC of the eukaryote; and (b) detecting specific hybridisation of detectably
 CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
 CC having a fragment identical to the predicted exon, the probe is included
 CC in the above mentioned microarray; assigning exons to a single gene,
 CC comprising (a) identifying exons from genomic sequence by the method
 CC above and (b) measuring the expression of each of the exons in several
 CC tissues and/or cell types using hybridisation to a single exon

CC microarrays having a probe with the exon, where a common pattern of
 CC expression of the exons in the tissues and/or cell types indicates that
 CC the exons should be assigned to a single gene; a peptide comprising one
 CC of 12011 sequences, mentioned in the specification, or encoded by the
 CC probes/open reading frames (ORF). The probes are used for gene
 CC expression analysis, and for identifying exons in a gene, particularly
 CC using human lung derived mRNA and for the study of lung diseases
 CC such as asthma, lung cancer, chronic obstructive pulmonary disease
 CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary
 CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,
 CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary
 CC haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis,
 CC pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension
 CC and hyaline membrane disease. The present sequence is a peptide/protein
 CC encoded by a single exon probe of the invention.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX Sequence 188 AA;
 QY
 Query Match 31.08; Score 801; DB 23; Length 188;
 Best Local Similarity 76.68; Pred. No. 7; 9e-65;
 Matches 151; Conservative 3; Mismatches 3; Indels 40; Gaps 2;
 QY 302 LANPGSGOIOIMOFLELSDSSNSCITWECTGFEKMTDPDVARWGRKSKPMNRY 361
 DB 12 LTLGSGOIQIMOPLELSDSSNSCITWEGTGEKMTDPDVARWGRKSKPMNRY 71
 QY 362 DKLSRALRYYYDKNIMTKVHPRESSMYKYPSSDLPYMSYHGKRRYKFDGIAQAQP- 420
 DB 72 DKLSRALRYYYDKNIMTKV-----HGKRRYKFDGIAQAQAQP 111
 QY 421 -----HAHPQKMFVAPHPALVYSSSFFPAAPNPYNSTGSIYP 461
 DB 112 PPESSLYKYPSSDLPYMSYNAHPQKMFVAPHPALVYSSSFFPAAPNPYNSTGSIYP 171
 QY 462 NTRLPAAHPSHIGTTY 478
 DB 172 NTRLPAAHPSHIGTTY 188

RESULT 15
 ABB31070
 ID ABB31070 standard; Peptide; 144 AA.
 XX
 AC ABB31070;
 XX
 DT 01-FEB-2002 (first entry)
 DE Peptide #3721 encoded by breast cell single exon nucleic acid probe.
 XX
 KW Human: microarray; single exon probe; gene expression; breast;
 KW disease; cancer.
 XX
 OS Homo sapiens.
 PN WO200157271-A2.
 XX
 PD 09-AUG-2001.
 XX
 PE 30-JAN-2001; 2001WO-US00662.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX

PA (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-496933/54.

DR
XX
XX
PT New spatially-addressable set of single exon nucleic acid probes,
PT useful for measuring gene expression in sample derived from human
PT breast, comprises number of single exon nucleic acid probes -

PS Claim 27; SEQ ID NO 14038; 327pp + sequence listing; English.

XX
CC The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human breast and Bt 474 cells. The method involves contacting
CC the probes with a collection of detectably labelled nucleic acids
CC derived from mRNA of human breast, and then measuring the label
CC bound to each probe of the microarray. The probes are useful for
CC verifying the expression of regions of genomic DNA predicted to
CC encode proteins. They are useful for gene discovery, and for
CC determining predisposition and/or prognosing breast disease. Gene
CC expression analysis is useful for assessing the toxicity of chemical
CC agents on cells. The microarray of this invention presents a far greater
CC diversity of probes for measuring gene expression, with far less bias
CC than expressed sequence tag microarrays. The method is suitable for
CC rapid production of functional information from genomic sequence. The
CC present sequence is a peptide encoded by a single exon nucleic acid
CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX
SQ Sequence 144 AA;

Query Match 25.0%; Score 647; DB 22; Length 144;

Best Local Similarity 75.0%; Pred. No. 6, 2e-51;

Matches 123; Conservative 1; Mismatches 0; Indels 40; Gaps 2;

QY 319 LLSDSNSNCITWEGTNGEFGKMTDPEVARRMGERKSKPMNMYDKLSRALRYYYDKNIMT 378

DB 1 LLSDSNSNSCITWEGTNGEFGKMTDPEVARRMGERKSKPMNMYDKLSRALRYYYDKNIMT 60

QY 379 KYHPRESSMKYKPSDLPYSSYHGKRYAKKDFHGIAQALP----- 420

DB 61 KV-----HGKRYAKKDFHGIAQALPAPPRESSLYKXPSDLPYMG 100

QY 421 --HAHPQKMNFYAHPALPYTSSSFFAAPNPMYNSPTGCIYPN 462

DB 101 SYHAHPQKMNFYAHPALPYTSSSFFAAPNPMYNSPTGCIYPN 144

Search completed: November 9, 2002, 16:50:23

Job time : 57.5985 secs

QY 421 HAHPOKMFVAPHPALPYTSSSFFAAPNPYNSPTGGIYPTNRLPAAHMPSHLGTY 478
 DB 421 HAHPOKMFVAPHPALPYTSSSFFAAPNPYNSPTGGIYPTNRLPAAHMPSHLGTY 478

RESULT 2

US-08-878-177-2
 ; Sequence 2, Application US/08878177
 ; Patent No. 6294354
 ; GENERAL INFORMATION:
 ; APPLICANT: Chugai Seiyaku Kabushiki Kaisha, 5001, Iwamoto et
 ; TITLE OF INVENTION: Cell Calcification Suppressing Proteins and Genes of
 ; TITLE OF INVENTION: the Proteins
 ; FILE REFERENCE: Chugai Seiyaku Kabushiki Kaisha 5001
 ; CURRENT APPLICATION NUMBER: US/08/878,177
 ; NUMBER OF SEQ ID NOS: 7
 ; SOFTWARE: Patentln Ver. 2.0
 ; SEQ ID NO 2
 ; LENGTH: 451
 ; TYPE: PRP
 ; ORGANISM: protein sequence from C-11 gene
 US-08-878-177-2

Query Match 87.4%; Score 2262.5; DB 4; Length 451;
 Best Local Similarity 86.3%; Pred. No. 9,9e-209;
 Matches 430; Conservative 0; Mismatches 1; Indels 67; Gaps 3;

QY 1 MASTIKREALSVSEDOSELECAVSGPHLAKTEKTAASSSEYGGTSKMSPRVPOQDMSLP 60
 DB 1 MASTIKREALSVSEDOSELECAVSGPHLAKTEKTAASSSEYGGTSKMSPRVPOQDMSLP 60
 QY 61 PARVTIKMECPNQVNGSRNSPDDCSVAGKGVSSSDVNGVNGSYMEKHLPPNMTT 120
 DB 61 PARVTIKMECPNQVNGSRNSPDDCSVAGKGVSSSDVNGVNGSYMEKHLPPNMTT 120
 QY 121 NERRVIVPADPTLMSTDHVRQMLEMAVKEYGLPDVDILLFONIDGKELCKMTKDDFORLT 180
 DB 121 NERRVIVPADPTLMSTDHVRQMLEMAVKEYGLPDVDILLFONIDGKELCKMTKDDFORLT 180
 QY 181 PSYNADILSHLYLRETPPLHLTSDVDKALQNSPRLMARNTGATFPPTSYYPEA 240
 DB 181 PSYNADILSHLYLRETPPLHLTSDVDKALQNSPRLMARNTGATFPPTSYYPEA 240
 QY 241 TORITTRPDLPEQARRSAMTSHSPTQSKATOPSSSTVPKTEDQRPQDLPYQIIGPTSS 300
 DB 241 TORITTRPDLPEQARRSAMTSHSPTQSKATOPSSSTVPKTEDQRPQDLPYQIIGPTSS 300
 QY 301 RLANPSSGOIQIOMQFLLELSDSSNSCITWEGTNGEFKMTDPDEVARRRMGERKSKPPNNA 360
 DB 301 RLANPSSGOIQIOMQFLLELSDSSNSCITWEGTNGEFKMTDPDEVARRRMGERKSKPPNNA 360
 QY 361 YOKLSALRYYYDKNIMTYV-----HPPSSMYTTPSLDPPMSSY 400
 DB 361 YOKLSALRYYYDKNIMTYV-----HPPSSMYTTPSLDPPMSSY 400
 QY 401 HGRKRYAKFDFHIGIAOALOPHAHPQKMFVAPHPALPYTSSSFFAAPNPYNSPTGGIY 460
 DB 401 HGRKRYAKFDFHIGIAOALOPHAHPQKMFVAPHPALPYTSSSFFAAPNPYNSPTGGIY 460
 QY 461 PNTRLPAAHMPSHLGTY 478
 DB 461 PNTRLPAAHMPSHLGTY 478
 QY 478 PNTRLPAAHMPSHLGTY 478
 DB 478 PNTRLPAAHMPSHLGTY 478

RESULT 3

US-08-343-443B-4
 ; Sequence 4, Application US/08343443B
 ; Patent No. 5968734
 ; GENERAL INFORMATION:
 ; APPLICANT: Aurias, Alain
 ; APPLICANT: Delattre, Olivier

APPLICANT: Desmaze, Chantal
 APPLICANT: Melot, Thomas
 APPLICANT: Peter, Martine
 APPLICANT: Ploouastel, Beatrice
 APPLICANT: Thomas, Gilles
 APPLICANT: Zucman, Jessica
 TITLE OF INVENTION: NUCLEIC ACID CORRESPONDING TO A GENE OF
 TITLE OF INVENTION: CHROMOSOME 22 INVOLVED IN RECURRENT CHROMOSOMAL
 TITLE OF INVENTION: TRANSLATIONS ASSOCIATED WITH THE DEVELOPMENT OF CANCEROUS
 TITLE OF INVENTION: TUMORS, AND NUCLEIC ACIDS OF FUSION RESULTING FROM SAID
 NUMBER OF SEQUENCES: 129
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Weiser & Associates
 STREET: 230 South Fifteenth Street
 CITY: Philadelphia
 STATE: PA
 COUNTRY: USA
 ZIP: 19102

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: AEDIT 1.0 DOS text editor
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/343,443B
 FILING DATE: 18-NOV-1994
 CLASSIFICATION: 514
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: PCT/FR93/00494
 FILING DATE: 19-MAY-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: FR 92/06123
 FILING DATE: 20-MAY-1992
 ATTORNEY/AGENT INFORMATION:
 NAME: Weiser, Gerard J.
 REGISTRATION NUMBER: 19,763
 REFERENCE/DOCKET NUMBER: 989.6121P
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 215-875-8383
 TELEFAX: 215-875-8394
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 452 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-343-443B-4

Query Match 57.3%; Score 1482; DB 2; Length 452;
 Best Local Similarity 57.7%; Pred. No. 8,8e-134;
 Matches 291; Conservative 55; Mismatches 80; Indels 78; Gaps 9;

QY 1 MASTIKREALSVSEDOSELECAVSGPHLAKTEKTAASSSEYGGTSKMSPRVPOQDMSLP 59
 DB 1 MASTIKREALSVSEDOSELECAVSGPHLAKTEKTAASSSEYGGTSKMSPRVPOQDMSLP 59
 QY 60 PARVTIKMECPNQVNGSRNSPDDCSVAGKGVSSSDVNGVNGSYMEKHLPPNMTT 118
 DB 60 PARVTIKMECPNQVNGSRNSPDDCSVAGKGVSSSDVNGVNGSYMEKHLPPNMTT 118
 QY 118 TINNERVIVPADPTLMSTDHVRQMLEMAVKEYGLPDVDILLFONIDGKELCKMTKDDFOR 178
 DB 118 TINNERVIVPADPTLMSTDHVRQMLEMAVKEYGLPDVDILLFONIDGKELCKMTKDDFOR 178
 QY 178 LTPSYNADILSHLYLRETPPLHLTSDVDKALQNSPRLMARNTGATFPPTSYYPEA 238
 DB 178 LTPSYNADILSHLYLRETPPLHLTSDVDKALQNSPRLMARNTGATFPPTSYYPEA 238
 QY 238 EATQRTTRPDLPEQARRSAMTSHSPTQSKATOPSSSTVPKTEDQRPQDLPYQIIGPTSS 297
 DB 238 EATQRTTRPDLPEQARRSAMTSHSPTQSKATOPSSSTVPKTEDQRPQDLPYQIIGPTSS 297
 QY 297 DOSSRLSVKEDSYDSVRKANGNNGNGLNKSPLGAGQITISKNTQRPQDLPYQIIGPTSS 268
 DB 297 DOSSRLSVKEDSYDSVRKANGNNGNGLNKSPLGAGQITISKNTQRPQDLPYQIIGPTSS 268

QY	298	TSSLAPGSSQIDLOMOFLLELLSDSSNCIMTEGNGEKKMTDPDEVARRMGERSKP	357
Db	269	TSSLAPGSSQIDLOMOFLLELLSDSNNAACITWEGTNGEKKMTDPDEVARRMGERSKP	328
QY	358	NMNTDKLSRALRYYYDNKINMTKHPPESSMYKYPDDLPMYSSYHGKKRAYKPFHGIQA	417
Db	329	NMNTDKLSRALRYYYDNKINMTKV-----HGKKRAYKPFHGIQA	368
QY	418	LQP-----HAHQKKNFVAHPHPALPTVSSSFFAAPNPMYNSPTG	457
Db	369	LQHPTESSMYKYPSPDISYMPSYHAHQOKNFVPPHPSSMPVTSSSFFGAASOYMTSPTG	428
QY	458	GIYPN---TRLPAHMHPSHGLTY	478
Db	429	GIYPNPVPRHPNTHVPSHGLSY	452

```

RESULT 4
US-09-360-779-2
; Sequence 2, Application US/09360779
; Patent No. 6268216
; GENERAL INFORMATION:
; APPLICANT: Deniers, Evan S.
; APPLICANT: Fyodorov, Dmitry V.
; APPLICANT: Hendricks, Timothy J.
; TITLE OF INVENTION: Reagents and Methods for the Screening of Compounds
; FILE REFERENCE: Case-03828
; CURRENT APPLICATION NUMBER: US/09/360,779
; CURRENT FILING DATE: 1999-07-26
; EARLIER APPLICATION NUMBER: 60/094,264
; EARLIER FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 340
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-09-360-779-2

Query Match          18.2%   Score 472;   DB 4;   Length 340;
Best Local Similarity 45.3%;   Pred. No. 5.3e-37;
Matches 105;   Conservative 13;   Mismatches 36;   Indels 78;   Gaps 7;

QY      296  GPTSSRLANPGSGQIQLMQFLLELLSDSSNSNCITWECTNGEFKMTDDEVARRWGERKS 355
          |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB      137  GPLSPAV-QKSGGQIQLMQFLLELLADBRANAGCIAMEGGHGFKLTDDDEVARRWGERKS 195

QY      356  KPNMNVOKLSALALYYDDKNTMTKVHPPESSMKYKPPSPLPTWSSIHGKRIYAKFFPHGIA 415
          |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB      196  KPNMNVOKLSALALYYDDKNTMTKVHPPESSMKYKPPSPLPTWSSIHGKRIYAKFFPHGIA 235

QY      416  QALQP---HAHP-----OQMNVAPHPPALPYTS 441
          |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB      236  QACGPPPAHANAAAAAAGDGAALYKLPAGLAPLPFPGLSLNLMAASAGVAPAGF 295

QY      442  SSFPAALNPYVNSP-----TGCITYNTKLP-----AAHMFSLGITY 478
          |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB      296  S-----YWPGPATATAAATAALYPTPGLAPPPPGFGVAAASHHGHHY 339

RESULT 5
US-09-435-335-2
; Sequence 2, Application US/09435335
; Patent No. 6384204
; GENERAL INFORMATION:
; APPLICANT: Deniers, Evan S.
; APPLICANT: Fyodorov, Dmitry V.
; APPLICANT: Hendricks, Timothy J.
; TITLE OF INVENTION: Reagents and Methods for the Screening of Compounds
; FILE REFERENCE: Case-04027
; CURRENT APPLICATION NUMBER: US/09/435,335

```

```

: CURRENT FILING DATE: 1999-11-05
: EARLIER APPLICATION NUMBER: 09/360,779
: EARLIER FILING DATE: 1999-07-26
: NUMBER OF SEQ ID NOS: 23
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 2
:   LENGTH: 340
:   TYPE: PRT
:   ORGANISM: Rattus norvegicus
US-09-435-335-2

Query Match      18.2%   Score 472; DB 4; Length 340;
Best Local Similarity 45.3%   Pred. No. 5.3e-37;
Matches 105; Conservative 13; Mismatches 36; Indels 78; Gaps 7

```

```

QY      286 CPTSLRANPGSGOIQLOMQLLELLSDSSNSKCTWEGTNGEFKMTDPEVARRMGERKS   355
        ||| :|||||||:|||||||:|||||||:|||||||:|||||||:|||||||:|||||||:
Db      137 GPLSPAV--OKSGSQIOLWOFLLELLADLANAGCIAMEGGHGEFKTLTDEPVARRMGERS   195
QY      356 KPNMYDLSTRRLRYVYDKNIMTKYHPPESSMKYKPSDLPYSSVHGKRYAKPDFHGIA   415
        ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      196 KPNMYDKLSRLRLRYTDKNIMSKY-----HGKRYARFPDQGLA   235
QY      416 QALOP---HAHP-----OKNFVAHPPALPYTS   441
        ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      236 OACOPPPAHAAHHAAAAAAAAAQQDALYKPLPAGLAPLPFPGISKLNTMAASAGVAPAQF   295
QY      442 SSFFAAPMPYNNSP-----TGCTTPRTLRP-----AAHMFSLGTTY   478
        ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      296 S-----YWPGPATATAAAAAATAALTALYPTPLGLOPPGPGGAVALAASHLGHY   339

RESULT 6
US-09-092-636--9
; Sequence 9, Application US/09092636A
; Patent No. 6162641
; GENERAL INFORMATION:
; APPLICANT: Goldman, Daniel
; APPLICANT: Sapru, Mohan K.
; TITLE OF INVENTION: Neuregulin Response Element and Uses Therefor
; FILE REFERENCE: UMI-003
; CURRENT APPLICATION NUMBER: US/09/092,636A
; CURRENT FILING DATE: 1998-06-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: patentin Ver. 2.0
; SEQ ID NO 9
; LENGTH: 468
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-092-636--9

Query Match          16.5%; Score 427; DB 4; Length 468;
Best Local Similarity 30.3%; Pred. NO. 1.9e-32;
Matches 115; Conservative 42; Mismatches 100; Indels 122; Gaps 12;

QY      122 ERRIVPADPLTMSTDHROWMLEMAYKEGYEDVDVILLFQNIDGKEIKMTKDDEQRLTP   181
        ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      94 QRRIGIPKNPMLMSGOVCOMLMTATNEFSLVNVNLHOF--GNNGOMLCWGERPLELAP   152
QY      182 SYNDILLSHLYLRE-----TPLPH-LTSDVDVKALONSPRLMHART   224
        ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      153 DFVGDIILWEHLQMIKENQEKTEDQYEENSHLNAPVHMINSITLGFSEGAQAYGMQAPR-   211
QY      225 GCATFIPIINTSVY----PEAT-----QRITT-----RPDLPEQAR--   256
        ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      212 ----YPKOWILDMSMCPSPATPAALGSELQMLPKSRILTNYVNVCYSISODFPSSNVLL   265
QY      257 -----RSAMTSSH-----PT-----QSK   270
        ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      266 NNNNGKPMDHDSPEKGGSFESSDSLRSNMNSOSSLLDYQRYPSFESEEDCSQSGLCSK   325
QY      271 ATQSSSTVPKTEDORPOLDPYQILGPTSSRLANPGSGOIOLMOFLLELLSDSSNSKCT   330
        ||| :|||||||:|||||||:|||||||:|||||||:|||||||:|||||||:|||||||:

```

[illegible]

```

1  NUMBER OF SEQUENCES: 4
2  CORRESPONDENCE ADDRESS:
3  ADDRESSEE: Incyte Pharmaceuticals, Inc.
4  STREET: 3174 Porter Dr.
5  CITY: Palo Alto
6  STATE: CA
7  COUNTRY: USA
8  ZIP: 94304
9  COMPUTER READABLE FORM:
10 MEDIUM TYPE: Diskette
11 COMPUTER: IBM Compatible
12 OPERATING SYSTEM: DOS
13 SOFTWARE: FastSeq for Windows Version 2.0
14 CURRENT APPLICATION DATA:
15 APPLICATION NUMBER: US/09/055,113
16 FILING DATE: Filed Herewith
17 PRIOR APPLICATION DATA:
18 APPLICATION NUMBER:
19 FILING DATE:
20 ATTORNEY/AGENT INFORMATION:
21 NAME: Billings, Lucy J.
22 REGISTRATION NUMBER: 36,749
23 REFERENCE/DOCKET NUMBER: PR-0501 US
24 TELECOMMUNICATION INFORMATION:
25 TELEPHONE: 650-855-0555
26 TELEFAX: 650-845-4166
27 TELEX:
28 INFORMATION FOR SEQ ID NO: 4:
29 SEQUENCE CHARACTERISTICS:
30 LENGTH: 452 amino acids
31 TYPE: amino acid
32 STRANDEDNESS: single
33 TOPOLOGY: linear
34 IMMEDIATE SOURCE:
35 LIBRARY: Genbank
36 CLONE: 511283
37 US-09-055-113-4
38
39 Query Match 12.2%, Score 317, DB 4, Length 452;
40 Best Local Similarity 24.5%, Pred. No. 6,4e-22;
41 Matches 106, Conservative 61, Mismatches 127, Indels 138, Gaps 17,
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960

```

RESULT 9
US-08-469-412A-7
; Sequence 7, Application US/08469412A
; Patent No. 5856125
; GENERAL INFORMATION:
; APPLICANT: Mavrothalassitis, George J.
; APPLICANT: Blair, Donald G.
; APPLICANT: Fisher, Robert J.
; APPLICANT: Beal Jr., Gregory J.
; APPLICANT: Athanasiou, Meropi A.
; APPLICANT: Sgouras, Dionyssios N.
; TITLE OF INVENTION: The ERF Genetic Locus and Its Products
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,412A
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Garrett-Mackowski, Eugenia
; REGISTRATION NUMBER: 37,330
; REFERENCE/DOCKET NUMBER: 015280-229000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 543 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..543
; OTHER INFORMATION: (first 8 amino acids from first exon not
; OTHER INFORMATION: included)
; US-08-469-412A-7

Query Match 11.7%; Score 303.5; DB 2; Length 543;
Best Local Similarity 39.2%; Pred. No. 1.7e-20;
Matches 69; Conservative 16; Mismatches 40; Indels 51; Gaps 4;

QY 303 ANPSSGQIQLMQFLELLSDSSNSNCITWEGTNGEFTKTDDEVARARRGERSKRNMYD 362
DB 12 SSPSSRQIQLMHFLELLRKEEYGVIAWGDYGEFVLDPEVARLMGVRCKCFQNMND 71
QY 363 KLSRALRYTYDKNIMTKVHPRESSMYKYPISDLPYWSSYHGKRYAKFDFH----- 412
DB 72 KLSRALRYTYDKNIMTKVHPRESSMYKYPISDLPYWSSYHGKRYAKFDFH----- 412
QY 413 --GIAQALQPHANP-----QKMNFAV-----HPALPYTSSSFFAA 447
DB 112 DMGLAGAVPQSAAPVPVSGSHFRFPSTPSEVLSPTEDPSPSPACSSSSSLFSA 167

RESULT 10
US-09-021-715-7
; Sequence 7, Application US/09021715

; Patent No. 6194547
; GENERAL INFORMATION:
; APPLICANT: Mavrothalassitis, George J.
; APPLICANT: Blair, Donald G.
; APPLICANT: Fisher, Robert J.
; APPLICANT: Beal Jr., Gregory J.
; APPLICANT: Athanasiou, Meropi A.
; APPLICANT: Sgouras, Dionyssios N.
; TITLE OF INVENTION: The ERF Genetic Locus and Its Products
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/021,715
; FILING DATE: 10-Feb-1998
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Garrett-Mackowski, Eugenia
; REGISTRATION NUMBER: 37,330
; REFERENCE/DOCKET NUMBER: 015280-229000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 543 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..543
; OTHER INFORMATION: (first 8 amino acids from first exon not
; OTHER INFORMATION: included)
; SEQUENCE DESCRIPTION: SEQ ID NO: 7:
; US-09-021-715-7

Query Match 11.7%; Score 303.5; DB 4; Length 543;
Best Local Similarity 39.2%; Pred. No. 1.7e-20;
Matches 69; Conservative 16; Mismatches 40; Indels 51; Gaps 4;

QY 303 ANPSSGQIQLMQFLELLSDSSNSNCITWEGTNGEFTKTDDEVARARRGERSKRNMYD 362
DB 12 SSPSSRQIQLMHFLELLRKEEYGVIAWGDYGEFVLDPEVARLMGVRCKCFQNMND 71
QY 363 KLSRALRYTYDKNIMTKVHPRESSMYKYPISDLPYWSSYHGKRYAKFDFH----- 412
DB 72 KLSRALRYTYDKNIMTKVHPRESSMYKYPISDLPYWSSYHGKRYAKFDFH----- 412
QY 413 --GIAQALQPHANP-----QKMNFAV-----HPALPYTSSSFFAA 447
DB 112 DMGLAGAVPQSAAPVPVSGSHFRFPSTPSEVLSPTEDPSPSPACSSSSSLFSA 167

RESULT 11
US-08-469-412A-2
; Sequence 2, Application US/08469412A
; Patent No. 5856125
; GENERAL INFORMATION:
; APPLICANT: Mavrothalassitis, George J.
; APPLICANT: Blair, Donald G.

APPLICANT: Fisher, Robert J.
APPLICANT: Beal Jr., Gregory J.
APPLICANT: Athanasiou, Meropi A.
APPLICANT: Sgouras, Dionysios N.
TITLE OF INVENTION: The ERF Genetic Locus and Its Products
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/469,412A
FILING DATE: 05-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Garrett-Wackowski, Eugenia
REGISTRATION NUMBER: 37,330
REFERENCE/DOCKET NUMBER: 015280-229000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 548 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-469-412A-2

Query Match 11.7%; Score 303.5; DB 2; Length 548;
Best Local Similarity 39.2%; Pred. No. 1.7e-20;
Matches 69; Conservative 16; Mismatches 40; Indels 51; Gaps 4;

QY 303 ANPGSGIOIOMFLEILSSNSNCITWEGTNGEFKMTDPDEVARWGRKSKPMNYD 362
DB 20 SSPGSRQIOIMHFIILRLKEEYQGYIAMQDYGEEFYIKDPDEVARIMGYRKCKPQMNYD 79
QY 363 KLSRALRYYYDKNIMTKVHPPESSMYKYPDDLPMSSYHGKRYAYKPDFH----- 412
DB 80 KLSRALRYYYNKRILHKT-----KGRRTYKFNFKLVNYPFI 119
QY 413 --GIAQALOPHAP-----QKMFVAP-----HPPALPVYSSSFFAA 447
DB 120 DVLGAGAVQSAAPVPSGSGHFRFPPTSPSEVLSPTEDRSPACSSSSSLFSA 175

RESULT 12
US-09-021-715-2
Sequence 2, Application US/09021715
Patent No. 6194347
GENERAL INFORMATION:
APPLICANT: Mavrothalassitis, George J.
Fisher, Robert J.
Beal Jr., Gregory J.
Athanasiou, Meropi A.
Sgouras, Dionysios N.
TITLE OF INVENTION: The ERF Genetic Locus and Its Products
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA

ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/021,715
FILING DATE: 10-Feb-1998
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Garrett-Wackowski, Eugenia
REGISTRATION NUMBER: 37,330
REFERENCE/DOCKET NUMBER: 015280-229000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 548 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-021-715-2

Query Match 11.7%; Score 303.5; DB 4; Length 548;
Best Local Similarity 39.2%; Pred. No. 1.7e-20;
Matches 69; Conservative 16; Mismatches 40; Indels 51; Gaps 4;

QY 303 ANPGSGIOIOMFLEILSSNSNCITWEGTNGEFKMTDPDEVARWGRKSKPMNYD 362
DB 20 SSPGSRQIOIMHFIILRLKEEYQGYIAMQDYGEEFYIKDPDEVARIMGYRKCKPQMNYD 79
QY 363 KLSRALRYYYDKNIMTKVHPPESSMYKYPDDLPMSSYHGKRYAYKPDFH----- 412
DB 80 KLSRALRYYYNKRILHKT-----KGRRTYKFNFKLVNYPFI 119
QY 413 --GIAQALOPHAP-----QKMFVAP-----HPPALPVYSSSFFAA 447
DB 120 DVLGAGAVQSAAPVPSGSGHFRFPPTSPSEVLSPTEDRSPACSSSSSLFSA 175

RESULT 13
US-08-343-443B-6
Sequence 6, Application US/0834443B
Patent No. 5968734
GENERAL INFORMATION:
APPLICANT: Aulias, Alain
APPLICANT: Delatize, Olivier
APPLICANT: Desmaze, Chantal
APPLICANT: Melot, Thomas
APPLICANT: Peter, Martine
APPLICANT: Ploougaestel, Beatrice
APPLICANT: Thomas, Gilles
APPLICANT: Zucman, Jessica
TITLE OF INVENTION: NUCLEIC ACID CORRESPONDING TO A GENE OF
TITLE OF INVENTION: CHROMOSOME 22 INVOLVED IN RECURRENT CHROMOSOMAL
TITLE OF INVENTION: TRANSLATIONS ASSOCIATED WITH THE DEVELOPMENT OF CANCEROUS
TITLE OF INVENTION: TUMORS, AND NUCLEIC ACIDS OF FUSION RESULTING FROM SAID
NUMBER OF SEQUENCES: 129
CORRESPONDENCE ADDRESS:
ADDRESSEE: Welsch & Associates
STREET: 230 South Fifteenth Street
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19102
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: AEDIT 1.0 DOS text editor
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/343.443B
FILING DATE: 18-NOV-1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/FR93/00494
FILING DATE: 19-MAY-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 92/06123
FILING DATE: 20-MAY-1992
ATTORNEY/AGENT INFORMATION:
NAME: Weiser, Gerard J.
REGISTRATION NUMBER: 19,763
REFERENCE/DOCKET NUMBER: 989.6121P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-875-8383
TELEFAX: 215-875-8394
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 109 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-343-443B-6

Query Match 11.6%; Score 301.5; DB 2; Length 109;
Best Local Similarity 65.5%; Pred. No. 2e-21;
Matches 57; Conservative 12; Mismatches 17; Indels 1; Gaps 1;

QY 252 YEQRASMTSHPTQSKAQ-PSSSTVPTEDQRPOLDPYQILGPTSSRLANPGSQI 310
DB 23 YDSVRGWMGNMNSGLKSPPLGAGQFISKNTEDRFPQDPYQILGPTSSRLANPGSQI 82
QY 311 QLMQFLLELSDSSNCTWEGTNGE 337
DB 83 QLMQFLLELSDSSNCTWEGTNGE 109

RESULT 14
US-09-055-113-1
Sequence 1, Application US/09055113
Patent No. 6265565
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Corley, Neil C.
APPLICANT: Guegler, Karl J.
APPLICANT: Lal, Preeti
TITLE OF INVENTION: PROSTATE-ASSOCIATED ETS PROTEIN
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Dr.
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/055.113
FILING DATE: Filed Herewith
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0501 US
TELECOMMUNICATION INFORMATION:

TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 335 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: PROSTUT12
CLONE: 1813005
US-09-055-113-1

Query Match 11.2%; Score 290.5; DB 4; Length 335;
Best Local Similarity 26.4%; Pred. No. 1.4e-19;
Matches 96; Conservative 41; Mismatches 114; Indels 113; Gaps 13;

QY 41 YGQTSKMSPRYPQDMLQPPARVITKMCNPQVNGSRNSPDDCSVAKGKMVSSSDNY 100
DB 67 YPEDSSMAKAKAPGASSREPP-----EEPEQCPIVDSQAPAGSLDLY 108
QY 101 GNNYSYMEKHI--PPNNMTNE-----RRVIYPADPTLSTDHVRQMLEAYKE 149
DB 109 P---GGLTLEHSLRQVQSMVYGEVLKDIETACKLLNTITADPMQSPSPVQKMLMTHERQ 165
QY 150 YGLPVDILLFQNDIGKELCKTKDFOFLAPSYNADILSLHYLRETPLPLHLSDDVD 209
DB 166 YRLPPMG-KAQQLAGKELCMSEQFQRSF-LGGDYVHAL-----DIW 209
QY 210 KALQNSPRIMHARNTGATFLFPNTSVYPEATQRTITRPDLPEQARRSAMTSHPTQS 269
DB 210 K---SAAMMKERTSPGAHYCASTS-----EESW----- 235
QY 270 KATOPSSSTVPTEDQRPQDLPYQILGPTSSRLANPGSQ-IQLMQFLLE-LLSSSNSN 327
DB 236 -----TDSVDSVSSCGQPIHLMQFLLELLKPKHSYGR 267
QY 328 CITW-EGTNGEFKMTDPDEVARRMGERSKPMNYDKLSRALRYYYDKNIMTKVHPRESS 386
DB 268 FIRMANKREKGFIEDSNQVARIWGIKRNPRAMNDKLSRSIROYYKKGIIRKPDLSQL 327
QY 387 MYKY 390
DB 328 YXQF 331

RESULT 15
US-08-469-412A-11
Sequence 11, Application US/08469412A
Patent No. 5856125
GENERAL INFORMATION:
APPLICANT: Mavrothalassitis, George J.
APPLICANT: Blair, Donald G.
APPLICANT: Fisher, Robert J.
APPLICANT: Beal Jr., Gregory J.
APPLICANT: Athanasios, Meropi A.
APPLICANT: Sgouras, Dionysios N.
TITLE OF INVENTION: The ERF Genetic Locus and Its Products
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/469,412A
FILING DATE: 05-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Garrett-Mackowski, Eugenia
REGISTRATION NUMBER: 37,330
REFERENCE/DOCKET NUMBER: 015280-229000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 78 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Domain
LOCATION: 1..78
OTHER INFORMATION: /note="ETS1 homologous region to
ets-like ERF DNA-binding domain"
US-08-469-412A-11

Query Match 10.9%; Score 283; DB 2; Length 78;
Best Local Similarity 55.1%; Pred. No. 7e-20;
Matches 54; Conservative 10; Mismatches 14; Indels 20; Gaps 1;
QY 312 LMQFLLELLSDSSNSNCITWEGTNGEFKMTDPDEVARRMGERSKPKPMNYDKLSRALRY 371
Db 1 LMQFLLELLDTRKSCQSFISWTGCGWEFKLSDPEVARRWGKRKNKPKMNYEKLSRGLRY 60
QY 372 YDKNIMTKVHPPESSMKYIPSDLPYMSYHGKRYAYK 409
Db 61 YDKNIITIKT-----AGKRYRYF 78

Search completed: November 9, 2002, 16:55:38
Job time : 24.6249 secs

TITLE OF INVENTION: Reagents and Methods for the Screening of Compounds
FILE REFERENCE: Case-03628
CURRENT APPLICATION NUMBER: US/09/850,799
CURRENT FILING DATE: 2001-05-08
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/360,779
PRIOR FILING DATE: EARLIER FILING DATE: 1999-07-26
NUMBER OF SEQ ID NOS: 13
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 2
LENGTH: 340
TYPE: PRT
ORGANISM: Rattus norvegicus
US-09-850-799-2

Query Match 18.2%; Score 472; DB 10; Length 340;
Best Local Similarity 45.3%; Pred. No. 4.3e-31;
Matches 105; Conservative 13; Mismatches 36; Indels 78; Gaps 7;

QY 296 GPTSSRLANPGSGOIQIWMQFLLELLSDSSNSNCITWBTGCEFFKMTDPDEVARRMGERKS 355
DB 137 GPLSPAV-QKSGGOIQIWMQFLLELLADLANAGCIAMEGHEGFLTPDEVARRMGERKS 195
QY 356 KPNMNYDKLSALRYYYDKNMTKVPHESSMYKPPSDLPYMSYHGKRYAKEDHGA 415
DB 196 KPNMNYDKLSALRYYYDKNMTKVPHESSMYKPPSDLPYMSYHGKRYAKEDHGA 235
QY 416 QALQP---HAHP-----OKMNVAPHPALPYTS 441
DB 236 QACQPPAHMAHA 295
QY 442 SSFPAAPNPNWSP-----TGTGTPNTRLP-----AAHMPSHLGTYY 478
DB 296 S-----YMPGPNATAAAATAATLYTPGLPPPGPFCAVAASHLGHY 339

RESULT 4
US-09-925-300-1360
Sequence 1360, Application US/09925300
Patent No. US20020151681A1
GENERAL INFORMATION:
APPLICANT: Craig Rosen,
APPLICANT: Steve Ruben,
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA101
CURRENT APPLICATION NUMBER: US/09/925,300
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05988
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 1890
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 1360
LENGTH: 344
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (2)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-300-1360

Query Match 13.4%; Score 348; DB 10; Length 344;
Best Local Similarity 29.2%; Pred. No. 5.6e-21;
Matches 96; Conservative 37; Mismatches 94; Indels 102; Gaps 10;

QY 163 IDEKELCKMTKDFQRLTPSTNADILSLHLYRE-----TPLPH-LTS 205
DB 9 MNGQMLCNLGEKRFLELAPFVGDILWEHLQMIKENQEKTEQYENSHLTSYPHWINS 68
QY 206 DDVVKALNSPRLMHAN--TGAT-----PIFNTSVSPRATQRT 245

DB 69 NTLGHTQAPYQOTQYPPKGLDSMCPASTSVLSSEDFQMFPSKLSVSVTYCS 128
QY 246 TRPDLPEQAR---RSAMTSHSPYQSKAT---QPSSTVPKTEQDQRPQLDPYQILG--- 296
DB 129 VSQDFPGSNLMLTNNSTPRPDHDSPEGNADSFESSDLSLWSNSSLQVQRPSPES 188
QY 297 -----PTSS-----RLANP-----GSGOIQIWMQFLLELL 320
DB 189 FEDDCOSGLCNKPTMSFKDYIOERSDPEQGPVIPAVALAGFTGSGPIQLMQLLELL 248
QY 321 SSSNSNCITWGTGCEFFKMTDPDEVARRMGERKS KPNMNYDKLSALRYYYDKNMTKRV 380
DB 249 SDKSCQSTISWTGDGWERKLADPDEVARRMGERKRNKPNMNYDKLSALRYYYDKNMTKRV 308
QY 381 HPPESSMYKPPSDLPYMSYHGKRYAKF 409
DB 309 -----SGKRYVYKF 317

RESULT 5
US-09-866-356-4
Sequence 4, Application US/09866356
Patent No. US20020098543A1
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
Guegler, Karl J.
Lal, Preethi
TITLE OF INVENTION: PROSTATE-ASSOCIATED ETS PROTEIN
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Dr.
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/866,356
FILING DATE: 29-May-2001
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/055,113
FILING DATE: 1998-04-03
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0501 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 452 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: Genbank
CLONE: 511283
SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-866-356-4

Query Match 12.2%; Score 317; DB 10; Length 452;
Best Local Similarity 24.5%; Pred. No. 2.8e-18;
Matches 106; Conservative 61; Mismatches 127; Indels 138; Gaps 17;

QY 81 SPDDCSYAKGKGVSSSDNMGMYGTYEKHIPP-N-MTNERVIVPA---DPTLMS 135

```
Db      4  TPAQCSIKQERISYTPESPVPSTAS-STPLHVPVPRALMEEDS:IRLPNLRLOPIYWS 62
Qy      136 TDHVRQMLEMAVKEYGILPDVILLFONIDGKELCKMTKDDFORLTPSYNADILLSHLYL 195
Db      63  RDDVAQMLKMAEWFSLRPDSNTFE-MNGKALLLTKEDEFRRSP-HSGDVIYELLQHT 120
Qy      196 RETPLPRL-----TSDDV-----DKALNSPR-----LMH 220
Db      121 LKQKRPRIFFSPFPHGNSIHFTQPEVILHQNHEEDNCVQTPRPSVDVNHNPTELLH 180
Qy      221 AARTGATLTFPTSYTPRQITRTP-----DLPEQAR----- 256
Db      181 RSRSP-----ITNHRPSDPDEQRLSPDLNMRILSPERAQGPPEHNNHESYPLS 236
Qy      257 -----RSAMTSSHPTQSKATOPSSSTVPKTEDQRPOLDPYQILGPTSSRLANP-----T 298
Db      237 VSPMENNHCPSASESHPKKSPRQESTRYL-----QLMPSPIHMLPLNHRHSVDK 288
Qy      299 SSRLANP-----SGQIQ-----LMQFLLEL 319
Db      289 QSRLSEDGLHRECKPYNLSHREDLAYNNHIMVSPPEEHAMPIGRADCRILMDYVYOL 348
Qy      320 LSDSSNSNCITWEGTNGE-FKMTDPDEVARRMGERSKPRMNTDKLSRLRYDDKNIMT 378
Db      349 LSDSRFENTIRWEDKSKIFRIYDPNGLARLWGNHKNRTMYEKMSRALRHYDKNLIIR 408
Qy      379 KVHPPESSMYKY 390
Db      409 K-EPGQRLLEFRF 419
```

```
RESULT 6
US-10-108-605-125
; Sequence 125, Application US/10108605
; Patent No. US20020160934A1
; GENERAL INFORMATION:
; APPLICANT: Broadus, Julie
; APPLICANT: Stam, Lynn
; APPLICANT: Bachmann, Jane
; APPLICANT: Kamdar, Kim
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES FROM DROSOPHILA MELANOGASTER THAT ENCODE
; FILE REFERENCE: 31133B
; CURRENT APPLICATION NUMBER: US/10/108,605
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: US 09/761,142
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/176,418
; NUMBER OF SEQ ID NOS: 361
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 125
; LENGTH: 623
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-10-108-605-125
```

Query Match 11.7%; Score 302; DB 9; Length 623;

Best Local Similarity 45.0%; Pred. No. 7, 4e-17;

Matches 68; Conservative 14; Mismatches 31; Indels 38; Gaps 4;

```
Qy      262 SSSHPTQSKATOPSSSTVPKTEDQRPOLDPYQILGPTSSRLANP-----GSGQIQLMQFLLE 318
Db      479 AHQHPAAYMSTL-----GLDK-GILGGYTTQGGVPCPTGSGPIQLMQFLLE 523
Qy      319 LSDSSNSNCITWEGTNGEFKMTDPDEVARRMGERSKPRMNTDKLSRLRYDDKNIMT 378
Db      524 LLDKTKQSFISWTGDEWFKLTDPEVARRMGIRKNRPMYEKLSRGLRYDDKNIIH 583
Qy      379 KVHPPESSMYKYPSDLPYMSYHGKRYAYKF 409
Db      584 KT-----AGKRYTYRF 594
```

```
RESULT 7
US-10-108-605-129
; Sequence 129, Application US/10108605
; Patent No. US20020160934A1
; GENERAL INFORMATION:
; APPLICANT: Broadus, Julie
; APPLICANT: Stam, Lynn
; APPLICANT: Bachmann, Jane
; APPLICANT: Kamdar, Kim
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES FROM DROSOPHILA MELANOGASTER THAT ENCODE
; FILE REFERENCE: 31133B
; CURRENT APPLICATION NUMBER: US/10/108,605
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: US 09/761,142
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/176,418
; NUMBER OF SEQ ID NOS: 361
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 129
; LENGTH: 623
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-10-108-605-129
```

Query Match 11.7%; Score 302; DB 9; Length 623;

Best Local Similarity 45.0%; Pred. No. 7, 4e-17;

Matches 68; Conservative 14; Mismatches 31; Indels 38; Gaps 4;

```
Qy      262 SSSHPTQSKATOPSSSTVPKTEDQRPOLDPYQILGPTSSRLANP-----GSGQIQLMQFLLE 318
Db      479 AHQHPAAYMSTL-----GLDK-GILGGYTTQGGVPCPTGSGPIQLMQFLLE 523
Qy      319 LSDSSNSNCITWEGTNGEFKMTDPDEVARRMGERSKPRMNTDKLSRLRYDDKNIMT 378
Db      524 LLDKTKQSFISWTGDEWFKLTDPEVARRMGIRKNRPMYEKLSRGLRYDDKNIIH 583
Qy      379 KVHPPESSMYKYPSDLPYMSYHGKRYAYKF 409
Db      584 KT-----AGKRYTYRF 594
```

```
RESULT 8
US-09-126-945B-2
; Sequence 2, Application US/09126945B
; Patent No. US20010010934A1
; GENERAL INFORMATION:
; APPLICANT: Liberman, Towla A.
; APPLICANT: Oettgen, Joerg P.
; APPLICANT: Kunsch, Charles A.
; APPLICANT: Endress, Gregory A.
; APPLICANT: Rosen, Craig A.
; TITLE OF INVENTION: Prostate Derived Ets Factor
; FILE REFERENCE: 1488.1090000
; CURRENT APPLICATION NUMBER: US/09/126,945B
; CURRENT FILING DATE: 1998-07-31
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 2
; LENGTH: 335
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-126-945B-2
```

Query Match 11.2%; Score 290.5; DB 10; Length 335;

Best Local Similarity 26.4%; Pred. No. 2, 6e-16;

Matches 96; Conservative 41; Mismatches 114; Indels 113; Gaps 13;

```
Qy      41 YGQTSMSRVRVQQDWLSQPPARVITKMECNPOVNGSRNSPDGCVATGAGKMYSSDNV 100
Db      117 YGQTSMSRVRVQQDWLSQPPARVITKMECNPOVNGSRNSPDGCVATGAGKMYSSDNV 100
```

Db 67 YPEDSSMAKAPGASSREPP-----EEEOQCPVIDSQAPAGSLDY 108
Qy 101 GMMTGSYMEKH-----PPNMTTNE-----RRVIYPADPTLMSTDHVRQMLEAVKE 149
Db 109 P---GGTLEHSELEQVSMVYGEVLKDIETACKLINTADPMQSPSVQKMLMTBEHQ 165
Qy 150 YGLPVDLILFONIDGKELCKMTKDDFORLTPSYNADILSLHLRLTRETPLRLTSDVD 209
Db 166 YRLPPMG-KATQELAGKELCMASEQFQORSF-LGGDVLHHL-----DIW 209
Qy 210 KALONSPRLMARNGTGATFIFPNTSVYPEATQRTTRPDLPEQARRSAMTSHSHPTQS 269
Db 210 K-----SAAMKERTSPGAIHYCASTS-----EESW-----235
Qy 270 KATOPSSSTVKTEDQRPQDLPYQILGPTSSRLANPGSQ-IQLMQFLLE-LLSOSSNSN 327
Db 236 -----TDSEVSSCSGQPIHLMQFLKELLKPHSYGR 267
Qy 328 CITW-EGTNGEFKMTDPEVARRMGERKSKPMNATDKLSRLRYYYDKNIMTKVHPRESS 386
Db 268 FIRMNKKRGIFKIEDSAQVARLWGIKRNPRAMNDKLSRSIROYYKKGIIRKPDISOQL 327
Qy 387 MYKY 390
Db 328 YVQF 331

RESULT 9
US-09-866-356-1
; Sequence 1, Application US/09866356
; Patent No. US20020098543A1
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; Corley, Neil C.
; Guegler, Karl J.
; Lal, Preeti
; TITLE OF INVENTION: PROSTATE-ASSOCIATED ETS PROTEIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION NUMBER: US/09/866,356
; APPLICATION NUMBER: US/09-04-03
; FILING DATE: 29-May-2001
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/055,113
; FILING DATE: 1998-04-03
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0501 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-845-0555
; TELEFAX: 650-845-4166
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 335 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: PROSTUT12
; CLONE: 1813005

SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-866-356-1

Query Match 11.2%; Score 290.5; DB 10; Length 335;
Best Local Similarity 26.4%; Pred. No. 2,6e-16;
Matches 96; Conservative 41; Mismatches 114; Indels 113; Gaps 13;

Qy 41 YGQTSKMSPRYPQODWLSOPPARVITKMECPNNOYNSRNSPDCSVAKGGMVSSSDNV 100
Db 67 YPEDSSMAKAPGASSREPP-----EEEOQCPVIDSQAPAGSLDY 108
Qy 101 GMMTGSYMEKH-----PPNMTTNE-----RRVIYPADPTLMSTDHVRQMLEAVKE 149
Db 109 P---GGTLEHSELEQVSMVYGEVLKDIETACKLINTADPMQSPSVQKMLMTBEHQ 165
Qy 150 YGLPVDLILFONIDGKELCKMTKDDFORLTPSYNADILSLHLRLTRETPLRLTSDVD 209
Db 166 YRLPPMG-KATQELAGKELCMASEQFQORSF-LGGDVLHHL-----DIW 209
Qy 210 KALONSPRLMARNGTGATFIFPNTSVYPEATQRTTRPDLPEQARRSAMTSHSHPTQS 269
Db 210 K-----SAAMKERTSPGAIHYCASTS-----EESW-----235
Qy 270 KATOPSSSTVKTEDQRPQDLPYQILGPTSSRLANPGSQ-IQLMQFLLE-LLSOSSNSN 327
Db 236 -----TDSEVSSCSGQPIHLMQFLKELLKPHSYGR 267
Qy 328 CITW-EGTNGEFKMTDPEVARRMGERKSKPMNATDKLSRLRYYYDKNIMTKVHPRESS 386
Db 268 FIRMNKKRGIFKIEDSAQVARLWGIKRNPRAMNDKLSRSIROYYKKGIIRKPDISOQL 327
Qy 387 MYKY 390
Db 328 YVQF 331

RESULT 10
US-09-925-297-716
; Sequence 716, Application US/09925297
; Patent No. US20020081659A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA105
; CURRENT APPLICATION NUMBER: US/09/925,297
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05989
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 928
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 716
; LENGTH: 203
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-925-297-716

Query Match 10.8%; Score 280.5; DB 10; Length 203;
Best Local Similarity 39.6%; Pred. No. 8.4e-16;
Matches 63; Conservative 21; Mismatches 44; Indels 31; Gaps 4;

Qy 257 RSAMTSHSHPTQSKATOP-----SSSTVPKTEEDQRPQDLPYQIL-GPTSSRLANPGSQI 310
Db 5 RGGYFSSSHGEGFSYKDRILYFDDICVYPERLEGKVKQEPIMYRGRPPYOR-----RSL 59
Qy 311 QLMQFLLELLSDSSNSNCITWGTNGEFKMTDPEVARRMGERKSKPMNATDKLSRLRY 370
Db 60 QLMQFLVTLDDPANAHFIATWGRMEFKLIEPEVARRMGIQKRNPRAMNDKLSRLRY 119
Qy 371 YYDKNIMTKVHPRESSMKYPSDLPYMSYHCKRAYKF 409
Db 120 YKGIKQKV-----AGERYVYKF 138

```
RESULT 11
US-09-864-761-40154
; Sequence 40154, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aecm1ca X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 40154
; LENGTH: 55
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AP001426.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.76
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2
; OTHER INFORMATION: SWISSPROT HIT: P11308, EVALUATE 7.00e-27
; OTHER INFORMATION: EST_HUMAN HIT: A0136709.1, EVALUATE 8.00e-26
US-09-864-761-40154
```

```
Query Match 10.8%; Score 279; DB 10; Length 55;
Best Local Similarity 96.4%; Pred. No. 1.7e-16;
```

```
Matches 53; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 7 EALSVSEDSQSLFECAYGAGSLAKTETMTASSSSRYGQTSKMSRPVPOQDWLSQPP 61
Db 1 EALSVSEDSQSLFECAYGAGTFLAKTETMTASSSSRYGQTSKMSRPVPOQDWLSQPP 55
;
RESULT 12
US-09-841-963A-2
; Sequence 2, Application US/09841963A
; Patent No. US20020081601A1
; GENERAL INFORMATION:
; APPLICANT: Watson, Dennis K.
; APPLICANT: Papas, Tula C. (Legal Representative)
; APPLICANT: Papas, Tula C. (Legal Representative)
; TITLE OF INVENTION: Methods and compositions for the diagnosis and treatment of ca
; TITLE OF INVENTION: based on transcription factor Ems2
; FILE REFERENCE: 10545-015-999
; CURRENT APPLICATION NUMBER: US/09/841,963A
; CURRENT FILING DATE: 2001-04-25
; PRIOR APPLICATION NUMBER: PCT/US99/27805
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: 06/109,850
; PRIOR FILING DATE: 1998-11-25
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 335
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-841-963A-2
;
Query Match 10.5%; Score 270.5; DB 10; Length 335;
Best Local Similarity 25.5%; Pred. No. 1.1e-14;
Matches 93; Conservative 43; Mismatches 115; Indels 113; Gaps 13;
QY 41 YGQTSKMSRPVPOQDWLSQPPARYTITKECNPMQVNSRNSPDDCSYAKGKRVSSSDNY 100
Db 67 YPDDSSWAAPAGASSREPP-----DQPCQCPVDSQAPAGTLDLY 108
QY 101 GMYGSYMEERKHT--PPNNNTNE-----RVYVADPPLTMSGDHYRQMLENAVKE 149
Db 109 P---GGILTEHSELEOVQSVNVEVLKDIETACKLNLITADPMQSPSNVQKWLMTHER 165
QY 150 YGLPDVDILLFQNIIDKELCKMTKDDFORLTSPYNADILSHLYRRETPPLPLTSDVD 209
Db 166 YRLPPNG-KAFQELAGELCAMSEQFQRSP-LGGDVLHAHL-----DIW 209
QY 210 KALQNSPRLMHANRTGATFTFPNTSYVPATQRIITRPDLPEQARRSAMTSHSHPQS 269
Db 210 K---SAAMWKERTSPGAIHYCASTS-----EESW----- 235
QY 270 KATQPSSTVPTKEDQRPQDLPQIILGPTSSRLANPGSGO-IOLMQFILE-LTSDSSNS 327
Db 236 -----IDSEVDSSCSGQPHLWQFLKLELKLPHSYGR 267
QY 328 CITW-EGTNGEFKMTDPDEYARRRGRKSKPNNAYDKLSRALYYVDKNIIMTVHPRESS 386
Db 268 FIMWLKKEGIFKIEDSAQYARLWGIKRNRPANNCOKLSINSQYKKGIIRKPDISQRL 327
QY 387 MRY 390
Db 328 VYOF 331
;
RESULT 13
US-09-920-300A-1789
; Sequence 1789, Application US/09920300A
; Patent No. US20020136728A1
; GENERAL INFORMATION:
; APPLICANT: King, Gordon E.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Xu, Jiaqun
```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 9, 2002, 16:46:47 ; Search time 31.3864 Seconds
(without alignments)
1464.080 Million cell updates/sec

Title: US-09-902-772-4

Perfect score: 2588
Sequence: 1 MASTIKELSVSEDSLEF.....ITPNTRLPAHMPSHLGTYY 478

Scoring table:
BLOSUM62
Gapop 10.0 , Capext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: PIR.73:*
2: PIR1:*
3: PIR2:*
4: PIR3:*
5: PIR4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2436	94.1	478	1	transcription fact
2	2324.5	89.8	486	1	transcription fact
3	1482	57.3	452	1	transforming prote
4	1474	57.0	452	1	transforming prote
5	1440.5	55.7	453	1	transforming prote
6	1253.5	48.4	272	2	transcription fact
7	1232.5	47.6	385	1	transcription fact
8	596	23.0	173	2	transcription fact
9	560.5	21.7	179	2	transcription fact
10	506.5	19.6	377	2	transcription fact
11	445	17.6	464	1	transcription fact
12	445	17.2	208	2	transcription fact
13	444	17.2	438	1	transcription fact
14	443.5	17.1	454	1	transcription fact
15	441.5	17.1	454	1	transcription fact
16	441	17.0	479	1	transcription fact
17	440.5	17.0	441	1	transcription fact
18	435.5	16.8	454	2	transcription fact
19	435	16.8	472	1	transcription fact
20	434.5	16.8	441	1	transcription fact
21	434.5	16.8	485	1	transcription fact
22	433.5	16.8	441	1	transcription fact
23	432.5	16.7	469	1	transcription fact
24	430.5	16.6	440	1	transcription fact
25	427	16.5	468	1	transcription fact
26	426	16.5	472	1	transcription fact
27	425.5	16.4	488	1	transcription fact
28	421	16.3	103	2	transcription fact
29	415.5	16.1	440	2	transcription fact

30	350.5	13.5	718	1	gene pointed prote
31	344.5	13.3	94	2	transcription fact
32	326	12.6	268	2	transcription fact
33	318	12.3	211	2	transforming prote
34	315.5	12.2	732	2	ETS domain protein
35	314.5	12.2	250	2	PR-1 protein - hum
36	303.5	11.7	548	2	ETS2 repressor fac
37	302	11.7	623	1	gene pointed prote
38	300.5	11.6	110	2	transcription fact
39	293	11.3	477	1	transcription fact
40	288	11.1	477	1	transcription fact
41	280.5	10.8	393	2	transcription fact
42	280.5	10.8	510	1	transcription fact
43	278.5	10.8	342	2	transcription fact
44	278.5	10.8	761	2	88K E-26-specific
45	266	10.3	555	1	transcription fact

ALIGNMENTS

RESULT 1	S60754	transcription factor erg - chicken
C:Species:	Gallus gallus (chicken)	
C:Date:	27-Apr-1996	#sequence, revision 23-May-1997 #text, change 16-Jul-1999
C:Accession:	S60754	
R:Jordanian, P.; Dewitte, F.; Desblens, X.; Stehelin, D.; Duterque-Coquillard, M.		
Mech. Dev. 50, 17-28, 1995		
A:Title:	Mesodermal expression of the chicken erg gene associated with precartilagino	
A:Reference number:	S60754; MUID:95329425; PMID:7605748	
A:Accession:	S60754	
A>Status:	preliminary	
A:Molecule type:	mRNA	
A:Residues:	1-478 <DHO>	
A:Cross-references:	EMBL:X77159; MID:9790439; PIDN:CA54404.1; PID:9790440	
C:Genetics:		
A:Gene:	erg	
C:Superfamily:	transcription factor erg; ets DNA-binding domain homology; ets RII reg	
C:Keywords:	DNA binding; phosphoprotein; proto-oncogene; transcription factor; transf	
F:119-193/Domain:	ets RII regulatory region homology <ETR>	
F:312-390/Domain:	ets DNA-binding domain homology <ETS>	
Query Match	94.1% Score 2436; DB 1; Length 478;	
Best Local Similarity	92.0%; Pred. No. 3.3e-164;	
Matches	458; Conservative 0; Mismatches 0; Indels 40; Gaps 2;	
QY	1	MASTIKELSVSEDSLEFECAYGSPHLAKTEMTASSSEYGGTSKMSRPVPODMLSQP 60
DB	1	MASTIKELSVSEDSLEFECAYGSPHLAKTEMTASSSEYGGTSKMSRPVPODMLSQP 60
QY	61	PARTYTKMECPNOVNGSRNSPDDCSYAKGKMYSSDNVGMNYSYWEKHIPPNNMTT 120
DB	61	PARTYTKMECPNOVNGSRNSPDDCSYAKGKMYSSDNVGMNYSYWEKHIPPNNMTT 120
QY	121	NERVIVPADEPTLWSTDHVQWLEMAVKEVGLPVDILLFONIDKELCKMTKDFORT 180
DB	121	NERVIVPADEPTLWSTDHVQWLEMAVKEVGLPVDILLFONIDKELCKMTKDFORT 180
QY	181	PSYAADILLSHLATREPLPHLTSDDYDKALQNSPRLMHARNRGATFTFPNTSVYPEA 240
DB	181	PSYAADILLSHLATREPLPHLTSDDYDKALQNSPRLMHARNRGATFTFPNTSVYPEA 240
QY	241	TORTITRPDLPEQARRSASMTSHSHPTQSKATOPSSSTVPKTEPORPOLDYOLIGPSS 300
DB	241	TORTITRPDLPEQARRSASMTSHSHPTQSKATOPSSSTVPKTEPORPOLDYOLIGPSS 300
QY	301	RLANPGSQIQLMQLLELLSDSSNSNCITWEGNGEFGKMTDPDEVARMRERKSKPMN 360
DB	301	RLANPGSQIQLMQLLELLSDSSNSNCITWEGNGEFGKMTDPDEVARMRERKSKPMN 360
QY	361	YDKISRLARYYYDKNIMTKV-----HPSSSMKKYSDLPYSSSY 400
DB	361	YDKISRLARYYYDKNIMTKV-----HPSSSMKKYSDLPYSSSY 400

```

Db 361 YDKLSALRKYRYDKNIMTKYHGKRYAKKDFHGIQAOLDPHPRESSMYKYPDLPWMSY 420
QY 401 HGKRYAKKDFHGIQAOLDPHPAHPQKMFVAPPALPYTSSSFFAAPNPYMSPTGIGY 460
Db 421 -----HAHPQKMFVAPPALPYTSSSFFAAPNPYMSPTGIGY 460
QY 461 PNTRLPAHMPHSLGTY 478
Db 461 PNTRLPAHMPHSLGTY 478

RESULT 2
TYHUEG
transforming protein erg-3 - human
N:contains: transforming protein erg-1; transforming protein erg-2
C:Species: Homo sapiens (man)
C:Date: 31-Mar-1989 #sequence_revision 30-May-1997 #text_change 16-Jul-1999
C:Accession: A94294; A94178; I58410; A28041; A29515
R: Rao, V.N.; Papas, T.S.; Shyam, E.; Reddy, P.
Science 237, 635-639, 1987
A:Title: erg, a human ets-related gene on chromosome 21: alternative splicing, polyadeny
A:Reference number: A94294; MUID:87263429; PMID:3299708
A:Accession: A94294
A:Molecule type: mRNA
A:Residues: 1-231,256-486 <R&A>
A:Cross-references: GB:M17254; NID:9182186; PIDN:AAA52398.1; PID:9182187
R:Reddy, E.S.P.; Rao, V.N.; Papas, T.S.
Proc. Natl. Acad. Sci. U.S.A. 84, 6131-6135, 1987
A:Title: The erg gene: a human gene related to the ets oncogene.
A:Reference number: A94178; MUID:87317608; PMID:3476934
A:Accession: A94178
A:Molecule type: mRNA
A:Residues: 100-231,256-486 <R&B>
A:Cross-references: GB:M21535; NID:9182182; PIDN:AAA35811.1; PID:9182185
R:Prasad, D.D.; Rao, V.N.; Lee, L.; Reddy, E.S.
Oncogene 9, 669-673, 1994
A:Title: Differentially spliced erg-3 product functions as a transcriptional activator.
A:Reference number: I58410; MUID:94119611; PMID:8290279
A:Accession: I58410
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 230-259 <R&S>
A:Cross-references: GB:S68130; NID:9544960; PIDN:AA829724.1; PID:9544961
C:Genetics:
A:Gene: GDB:ERG
A:Cross-references: GDB:119884; OMIM:165080
A:Map position: 21q22.2-21q22.2
C:Superfamily: transcription factor erg; ets DNA-binding domain homology; ets RII regula
C:Keywords: alternative splicing; DNA binding; nucleus; proto-oncogene; transcription fa
F:1-231,256-486/Product: transforming protein erg-2 #status predicted <EG2>
F:100-231,256-486/Product: transforming protein erg-1 #status predicted <EG1>
F:126-200/Domain: ets RII regulatory region homology <ETR>
F:320-398/Domain: ets DNA-binding domain homology <ETS>

Query Match 89.8%; Score 2324.5; DB 1; Length 486;
Best Local Similarity 87.6%; Pred. No. 2.5e-156;
Matches 436; Conservative 9; Mismatches 12; Indels 41; Gaps 3;

```

```

QY 242 QRTTRPDLPEEQARRSAMTSHSPT-QSKATOPSSSTVPTKEDORPDLPPYQILGPTSS 300
Db 249 QRTTRPDLPEEQARRSAMTSHSPT-QSKATOPSSSTVPTKEDORPDLPPYQILGPTSS 308
QY 301 RLANPGSGOIQIOWOFLLELSDSSNSCTWECTNEEFKMTDDEVARRRGERSKRPNN 360
Db 309 RLANPGSGOIQIOWOFLLELSDSSNSCTWECTNEEFKMTDDEVARRRGERSKRPNN 368
QY 361 YDKLSALRKYRYDKNIMTKYHGKRYAKKDFHGIQAOLDPHPRESSMYKYPDLPWMSY 420
Db 369 YDKLSALRKYRYDKNIMTKYHGKRYAKKDFHGIQAOLDPHPRESSMYKYPDLPWMSY 408
QY 421 -----HAHPQKMFVAPPALPYTSSSFFAAPNPYMSPTGIGY 460
Db 409 HPESSLYKYPDLPYMSYHAHPQKMFVAPPALPYTSSSFFAAPNPYMSPTGIGY 468
QY 461 PNTRLPAHMPHSLGTY 478
Db 469 PNTRLPTSHMPHSLGTY 486

RESULT 3
137565
transforming protein flt, long splice form - human
N:Alternate names: Friend leukemia integration protein 1; transcription factor ERGB
C:Species: Homo sapiens (man)
C:Date: 04-Oct-1996 #sequence_revision 30-May-1997 #text_change 03-Mar-2000
C:Accession: I37565; S29843; S35506; A49000; A49015; I54170
R:Delattre, O.; Zucman, J.; Plougastel, B.; Desmaze, C.; Melot, T.; Peter, M.; Kovar,
Nature 359, 162-165, 1992
A:Title: Gene fusion with an ETS DNA-binding domain caused by chromosome translocatio
A:Reference number: S29843; MUID:9316799; PMID:8439553
A:Accession: S29843
A:Molecule type: mRNA
A:Residues: 1-452 <D&E>
A:Cross-references: EMBL:X67001; NID:932529; PIDN:CAA47399.1; PID:932530
R:Hromas, R.; May, W.; Denny, C.; Raskind, W.; Moore, J.; Maki, R.A.; Beck, E.; Klems
Biochim. Biophys. Acta 1172, 155-158, 1993
A:Title: Human FLI-1 localizes to chromosome 11Q24 and has an aberrant transcript in
A:Reference number: S29843; MUID:9316799; PMID:8439553
A:Accession: S29843
A:Molecule type: mRNA
A:Residues: 1-68, 'V', '70-76, '70-129, 'A', '131-132, 'V', '134-322, 'Q', '324-425, 427-452 <HRO1>
A:Cross-references: EMBL:M93255; NID:9182659
A:Note: the authors translated the codon GTC for residue 69 as Glu
R:Hromas, R.A.; May, W.; Denny, C.; Raskind, W.; Moore, J.; Maki, R.A.; Beck, E.; Kle
submitted to the EMBL Data Library, May 1993
A:Description: Human FLI-1, an ETS oncogene family member preferentially expressed in
A:Reference number: S35506
A:Accession: S35506
A:Molecule type: mRNA
A:Residues: 1-68, 'V', '70-76, '70-129, 'A', '131-132, 'V', '134-322, 'Q', '324-425, 427-452 <HRO2>
A:Cross-references: EMBL:M93255; NID:9182659; PIDN:AAA54479.1; PID:9182660
R:Prasad, D.D.; Rao, V.N.; Reddy, E.S.
Cancer Res. 52, 5833-5837, 1992
A:Title: Structure and expression of human FLI-1 gene.
A:Reference number: A49000; MUID:93007976; PMID:1394211
A:Accession: A49000
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-322, 'Q', '324-452 <P&A>
A:Cross-references: GB:S45205; NID:9257353; PIDN:AA823637.1; PID:9257354
A:Note: sequence extracted from NCBI backbone (NCBIN:115336, NCBI:115337)
R:Watson, D.K.; Smyth, F.E.; Thompson, D.M.; Cheng, J.O.; Testa, J.R.; Papas, T.S.; S
Cell Growth Differ. 3, 705-713, 1992
A:Title: The ERGB/FLI-1 gene: Isolation and characterization of a new member of the f
A:Reference number: A49015; MUID:93075640; PMID:1445800
A:Accession: A49015
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-390, 'Q', '392-425, 427-452 <WAT>

```

A:Cross-references: GB:M98933; NID:q182188; PIDN:AAA5812.1; PID:q182189
A:Note: sequence extracted from NCBI backbone (NCBI:119390, NCBI:119392)
A:Bad, V.; Lipinski, M.; Rassart, E.; Poliquin, L.; Bergeron, D.
Genomics 11, 223-224, 1991
A:Title: The human homolog of the mouse common viral integration region, FL11, maps to 1
A:Reference number: IS4170; MUID:92112219; PMID:1165382
A:Accession: IS4170
A>Status: preliminary; nucleic acid sequence not shown; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-6 <BAU>
A:Cross-references: GB:L47616; NID:g1000863; PIDN:AA476854.1; PID:g1000864
A:Genetics:
A:Gene: GDB:FL11
A:Cross-references: GDB:127565; OMIM:193067
A:Map position: 11q24.1-11q24.3
A:Insertions: 77/3
A:Superfamily: transcription factor erg; ets DNA-binding domain homology; ets RII reg
C:Keywords: alternative splicing; DNA binding; nucleus; proto-oncogene; transcription fa
F:118-192/Domain: ets RII regulatory region homology <ETR>
F:283-361/Domain: ets DNA-binding domain homology <ETS>

Query Match 57.3%; Score 1482; DB 1; Length 452;
Best Local Similarity 57.7%; Pred. No. 6e-97;

Matches 291; Conservative 55; Mismatches 80; Indels 78; Gaps 9;

QY 1 MASTIKKALSVSDQSLFECAYG-SPLAKTEMTASSSEYGTGTSKSPRPVPOQDWLSQ 59
1 MDGTRKALSVSDQSLFDSAYGAHAHLPRADMTASGSPYGGPHKINLPPOQEWLNQ 60
DB 1 MASTIKKALSVSDQSLFDSAYGAHAHLPRADMTASGSPYGGPHKINLPPOQEWLNQ 60
QY 60 PPAVYTIKMECPNOVNGSRNSPDDCSYAKGKMYSSSDNGMYGSMYEEKH-IPPPNM 118
1 PPAVYTIKMECPNOVNGSRNSPDDCSYAKGKMYSSSDNGMYGSMYEEKH-IPPPNM 118
DB 61 -PVAIVNKKREY--DHMGSRRESPPVDCSVSKCNKLVGGESNPMYNSYMDKNGPPEPPNM 117
1 PPAVYTIKMECPNOVNGSRNSPDDCSYAKGKMYSSSDNGMYGSMYEEKH-IPPPNM 118
QY 119 TTNERRYIVPADPTLMQSTIDHVRQMLEVAVKYGLPDVDILLFQNDKELCKMTKDFOR 178
1 TTNERRYIVPADPTLMQSTIDHVRQMLEVAVKYGLPDVDILLFQNDKELCKMTKDFOR 178
DB 118 TTNERRYIVPADPTLMQSTIDHVRQMLEVAVKYGLPDVDILLFQNDKELCKMTKDFOR 178
QY 179 LTPSYNDILSLHLYRETPPLHLTSDVDKALQNSPRLMHANNTGATFIPTTSYYP 238
1 LTPSYNDILSLHLYRETPPLHLTSDVDKALQNSPRLMHANNTGATFIPTTSYYP 238
DB 178 ATLYNTEVLLSHLYRESL-----LAVNTTSHT 208
QY 239 EATGRITTRPDLPEQARRSAMTSHPTQSKATQ-FSSSTVPTKTEQORPOLDPYQILGP 297
1 EATGRITTRPDLPEQARRSAMTSHPTQSKATQ-FSSSTVPTKTEQORPOLDPYQILGP 297
DB 209 DQSRRLVKKEDPSYDVSRGAMNNMNSGLKSPGLGAQISIKNTQRPDPDYQILGP 268
QY 298 TSSRLANPGSQIOLMQLFLELSDSSNSNCITWEGTNGEKKMTDPDEVARRMGERKSKP 357
1 TSSRLANPGSQIOLMQLFLELSDSSNSNCITWEGTNGEKKMTDPDEVARRMGERKSKP 357
DB 269 TSSRLANPGSQIOLMQLFLELSDSSNSNCITWEGTNGEKKMTDPDEVARRMGERKSKP 328
QY 358 NMNYDKLSRALRYYYDKNIMTKVHPRESSMYKYPDDLPMSSYHGKRAYKFDHGIQAQ 417
1 NMNYDKLSRALRYYYDKNIMTKVHPRESSMYKYPDDLPMSSYHGKRAYKFDHGIQAQ 417
DB 329 NMNYDKLSRALRYYYDKNIMTKV-----HGKRAYKFDHGIQAQ 368
QY 418 LQP-----HAHPQKMFVAPHPALPYTSSSFFAAPNPYWNSTPTG 457
1 LQP-----HAHPQKMFVAPHPALPYTSSSFFAAPNPYWNSTPTG 457
DB 369 LQHPRESSMYKYPDSISYMPSTYHAHQKVFVPHSSMPVITSSFFGAASQYWTSPGTG 428
QY 458 GIYPN---TRLPAAMPSHLGTYT 478
1 GIYPN---TRLPAAMPSHLGTYT 478
DB 429 GIYPNPNVPRHNPHTVPSHLGSY 452

RESULT 4

S17403
transforming protein fl1 - mouse

N:Alternate names: Friend leukemia integration protein 1; transcription factor ERGB
C:Species: Mus musculus (house mouse)
C>Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 16-Jul-1999
C:Accession: S17403

R:Ben-David, Y.; Giddens, E.B.; Letwin, K.; Bernstein, A.
Genes Dev. 5, 908-918, 1991
A:Title: Erythroleukemia induction by Friend murine leukemia virus: insertional activation
A:Reference number: S17403; MUID:91253758; PMID:2044959

A:Accession: S17403
A:Molecule type: mRNA
A:Residues: 1-452 <BEN>
A:Cross-references: GB:X59421; NID:g50974; PIDN:CAA42055.1; PID:g50975
A:Genetics:
A:Gene: FL1
C:Superfamily: transcription factor erg; ets DNA-binding domain homology; ets RII reg
C:Keywords: DNA binding; nucleus; proto-oncogene; transcription factor; transforming
F:118-192/Domain: ets RII regulatory region homology <ETR>
F:283-361/Domain: ets DNA-binding domain homology <ETS>

Query Match 57.0%; Score 1474; DB 1; Length 452;
Best Local Similarity 57.7%; Pred. No. 2.2e-96;

Matches 291; Conservative 53; Mismatches 82; Indels 78; Gaps 9;

QY 1 MASTIKKALSVSDQSLFECAYG-SPLAKTEMTASSSEYGTGTSKSPRPVPOQDWLSQ 59
1 MDGTRKALSVSDQSLFDSAYGAHAHLPRADMTASGSPYGGPHKINLPPOQEWLNQ 60
DB 1 MASTIKKALSVSDQSLFDSAYGAHAHLPRADMTASGSPYGGPHKINLPPOQEWLNQ 60
QY 60 PPAVYTIKMECPNOVNGSRNSPDDCSYAKGKMYSSSDNGMYGSMYEEKH-IPPPNM 118
1 PPAVYTIKMECPNOVNGSRNSPDDCSYAKGKMYSSSDNGMYGSMYEEKH-IPPPNM 118
DB 61 -PVAIVNKKREY--DHMGSRRESPPVDCSVSKCNKLVGGESNPMYNSYMDKNGPPEPPNM 117
1 PPAVYTIKMECPNOVNGSRNSPDDCSYAKGKMYSSSDNGMYGSMYEEKH-IPPPNM 118
QY 119 TTNERRYIVPADPTLMQSTIDHVRQMLEVAVKYGLPDVDILLFQNDKELCKMTKDFOR 178
1 TTNERRYIVPADPTLMQSTIDHVRQMLEVAVKYGLPDVDILLFQNDKELCKMTKDFOR 178
DB 118 TTNERRYIVPADPTLMQSTIDHVRQMLEVAVKYGLPDVDILLFQNDKELCKMTKDFOR 178
QY 179 LTPSYNDILSLHLYRETPPLHLTSDVDKALQNSPRLMHANNTGATFIPTTSYYP 238
1 LTPSYNDILSLHLYRETPPLHLTSDVDKALQNSPRLMHANNTGATFIPTTSYYP 238
DB 178 ATLYNTEVLLSHLYRESL-----LAVNTTSHT 208
QY 239 EATGRITTRPDLPEQARRSAMTSHPTQSKA-TQSSSTVPTKTEQORPOLDPYQILGP 297
1 EATGRITTRPDLPEQARRSAMTSHPTQSKA-TQSSSTVPTKTEQORPOLDPYQILGP 297
DB 209 DQSRRLVKKEDPSYDVSRGAMNNMNSGLKSPGLGSQYMGKNTQRPDPDYQILGP 268
QY 298 TSSRLANPGSQIOLMQLFLELSDSSNSNCITWEGTNGEKKMTDPDEVARRMGERKSKP 357
1 TSSRLANPGSQIOLMQLFLELSDSSNSNCITWEGTNGEKKMTDPDEVARRMGERKSKP 357
DB 269 TSSRLANPGSQIOLMQLFLELSDSSNSNCITWEGTNGEKKMTDPDEVARRMGERKSKP 328
QY 358 NMNYDKLSRALRYYYDKNIMTKVHPRESSMYKYPDDLPMSSYHGKRAYKFDHGIQAQ 417
1 NMNYDKLSRALRYYYDKNIMTKVHPRESSMYKYPDDLPMSSYHGKRAYKFDHGIQAQ 417
DB 329 NMNYDKLSRALRYYYDKNIMTKV-----HGKRAYKFDHGIQAQ 368
QY 418 LQP-----HAHPQKMFVAPHPALPYTSSSFFAAPNPYWNSTPTG 457
1 LQP-----HAHPQKMFVAPHPALPYTSSSFFAAPNPYWNSTPTG 457
DB 369 LQHPRESSMYKYPDSISYMPSTYHAHQKVFVSHPSMPVITSSFFGAASQYWTSPGT 428
QY 458 GIYPN---TRLPAAMPSHLGTYT 478
1 GIYPN---TRLPAAMPSHLGTYT 478
DB 429 GIYPNPNVPRHNPHTVPSHLGSY 452

RESULT 5

S49013
transforming protein fl1 - African clawed frog

N:Alternate names: Friend leukemia integration protein 1; transcription factor ERGB
C:Species: Xenopus laevis (African clawed frog)
C>Date: 07-May-1995 #sequence_revision 23-May-1997 #text_change 16-Jul-1999
C:Accession: S49013

R:Meier, D.; Wolff, C.M.; Stiegler, P.; Senan, F.; Befort, N.; Befort, J.J.; Remy, P.
Mech. Dev. 44, 109-121, 1993
A:Title: X1-fl1, the Xenopus homologue of the fl1-1 gene, is expressed during embryog

A:Reference number: S49013; MUID:94206844; PMID:8155576

A:Accession: S49013

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-453 <MEY>

A:Cross-references: EMBL:X66979; NID:g505486; PIDN:CAA47389.1; PID:g505487

C:Genetics:

A:Gene: fl1

C:Superfamily: transcription factor erg; ets DNA-binding domain homology; ets RII reg
C:Keywords: DNA binding; nucleus; proto-oncogene; transcription factor; transforming
F:117-191/Domain: ets RII regulatory region homology <ETR>

Db 321 YMPSTHAHQVNFVPPHPSMPVTSSTSFQASQYWT-S-IGITPNDNVRHPTHTVPS 379
QY 473 HLGYTY 478
|||:|
Db 380 HLGSYTY 385

RESULT 8

A56646
transcription factor erg/fli-1 homolog - sea urchin (Lytechinus variegatus) (fragment)
C:Species: Lytechinus variegatus (variegated urchin)
C>Date: 03-Oct-1995 #sequence_revision 03-Oct-1995 #text_change 16-Jul-1999
C:Accession: A56646
R:Q1, S.: Chen, Z.Q.; Papas, T.S.; Lautenberger, J.A.
DNA Seq. 3, 127-130, 1992
A:Title: The sea urchin erg homolog defines a highly conserved erg-specific domain.
A:Reference number: A56646; MUID:93091246; PMID:1457815
A:Accession: A56646
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-173 <Q1A>
A:Cross-references: GB:M81067; NID:g161310; PIDN:AAA68905.1; PID:g161311
C:Superfamily: transcription factor erg; ets DNA-binding domain homology; ets RII regula
C:Keywords: DNA binding; nucleus; transcription factor
F:6-84/Domain: ets DNA-binding domain homology <ETS>

Query Match 23.0%; Score 596; DB 2; Length 173;
Best Local Similarity 61.9%; Pred. No. 4.8e-35;
Matches 120; Conservative 13; Mismatches 17; Indels 44; Gaps 5;

QY 307 SGOIQLOMOLFLELSDSSNSCITWEGTNGEFKMTDPEVARRMGERKSPNNMYDKLSR 366
|||||
Db 1 SGOIQLOMOLFLELSDSSNSCITWEGTNGEFKMTDPEVARRMGERKSPNNMYDKLSR 60

QY 367 ALRYVYKNIWTKYHPPSSMYKKYPSDLRYMSYHGRKYAKKPEHGIQAOLQ----- 419
|||||
Db 61 ALRYVYKNIWTKY-----HGKRYAKKFPAGIAQAMQVQADPS 100

QY 420 -----PHAHPOKMFVAPHPALPYTSSSFPAAPRYMNSPG-CIYPN---TR 464
|||:|
Db 101 MYRYQSDLTLYPGYHPTKLNFGV--TRINPTNMSLPSHSHSYSSPITGANTYPSGHYTH 158

QY 465 LPAAHMPSHLGTY 478
|||:|
Db 159 PHASHMSHIGTY 172

RESULT 9

S51226
transcription factor erg/fli-1 homolog - polychaete (Nereis diversicolor) (fragment)
C:Species: Nereis diversicolor (sandworm)
C>Date: 15-Jul-1995 #sequence_revision 23-May-1997 #text_change 30-May-1997
C:Accession: S51226
R:Leleuvre-Chotteau, A.; Laudet, V.; Flourens, A.; Begue, A.; Leprince, D.; Fontaine, F.
FEBS Lett. 354, 62-66, 1994
A:Title: Identification of two ets related genes in a marine worm, the polychaete annelid
A:Reference number: S51226; MUID:95046307; PMID:7957902
A:Accession: S51226
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-179 <TEL>
C:Superfamily: transcription factor erg; ets DNA-binding domain homology; ets RII regula
C:Keywords: DNA binding; nucleus; transcription factor
F:7-85/Domain: ets DNA-binding domain homology <ETS>

Query Match 21.7%; Score 560.5; DB 2; Length 179;
Best Local Similarity 57.4%; Pred. No. 1.6e-32;
Matches 113; Conservative 14; Mismatches 25; Indels 45; Gaps 5;

QY 306 GSGOIQLOMOLFLELSDSSNSCITWEGTNGEFKMTDPEVARRMGERKSPNNMYDKLS 365
|||||
Db 1 GSGOIQLOMOLFLELSDSSNSCITWEGTNGEFKMTDPEVARRMGERKSPNNMYDKLS 60

QY 366 RALRYTYKNIWTKYHPPSSMYKTPSDLPYMSYHGRKYAKKPEHGIQAOLQ----- 420
|||||
Db 61 RALRYTYKNIWTKY-----HGKRYAKKFPAGIAQAMQVSTTDP 100

QY 421 -----HAHPQKMFVAPHPALPYTSSSFPAAPRYMNSPG-CIYNTPLRP 466
|||:|
Db 101 AAYKYQODLMSGYHHTSKNLMAHAP-MASSASGFPFPAPRYMSLVGSLNLYPNISNH 159

QY 467 A-----AHMPSHLGTY 478
|||:|
Db 160 AMSHHGMSHLSYTY 176

RESULT 10
T34343
hypothetical protein T08H4.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 02-Sep-2000
C:Accession: T34343
R:Stelljes, L.
submitted to the EMBL Data Library, October 1995
A:Description: The sequence of C. elegans cosmid T08H4.
A:Reference number: Z21510

A:Accession: T34343
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-377 <STEE>
A:Cross-references: EMBL:U039470; PIDN:AACT1120.1; GSPDB:GN00020; CESP:T08H4.3
A:Experimental source: strain Bristol N2; clone T08H4
C:Genetics:

A:Gene: CESP:T08H4.3
A:Map position: 2
A:Insertions: 63/73; 88/1; 111/73; 146/1; 194/1; 297/1; 343/3
F:216-294/Domain: ets DNA-binding domain homology <ETS>

Query Match 19.6%; Score 506.5; DB 2; Length 377;
Best Local Similarity 38.9%; Pred. No. 3e-28;
Matches 133; Conservative 26; Mismatches 76; Indels 107; Gaps 13;

QY 154 DVDILLFQNIQD--GKECLKMKRDKDQRLTPSYNADILLSHLYLRETPHLTSDV--D 209
|||:|
Db 27 DEDVTKYSAIQTIKQEQOQOQOQSNALPSN-----FFFNQMOMDFPN 73

QY 210 KALQNSPRLMARNTGAGATFIFPNTSVYPEATQITTRPDLRYEQARSAWTSRSH----- 265
|||:|
Db 74 RMLYNDNTMOKSEND--HFTGMNLSTASSSGNGNSTSKD-----QSRQFTYESNSNGN 126

QY 266 -----PTQSR-----ATQPSSTVPTKEDQF-POL----- 289
|||:|
Db 127 GAATSGSGSSSSPESKSDYFNISMNAFAATPGS-----KSPDHNPFSFNMLSYTGAL 182

QY 290 -----DPRYITLPTSSRLANPSSGOIQLOMOLFLELSDSSNSCITWEGTNGEF 338
|||||
Db 183 KLSNSTSPANDPYYOILPPTSKNLASHSSGQIQLOMOLFLELSDRYEVIWTEGTQEF 242

QY 339 KMTDPEVARRMGERKSPNNMYDKLSRALRYRYKNIWTKYHPPSSMYKKYPSDLRYMS 398
|||:|
Db 243 KLVDPDEVARRMGERKSPNNMYDKMSRALRYRYKNIWTKY----- 284

QY 399 SYHGRKYAKKPEHGIQAOLQ--HAHPQ-----KMFVAP 432
|||||
Db 285 --HGKRYAKKPEHGIQAOLQDPPTASHQDYFNSHAMGRIVP 324

RESULT 11
S37616
transcription factor elg - fruit fly (Drosophila melanogaster)
N:Alternate names: transcription factor ets-97D
C:Species: Drosophila melanogaster
C>Date: 13-Jan-1995 #sequence_revision 23-May-1997 #text_change 16-Jul-1999
C:Accession: S37616; S28822
R:The, S.M.; Xie, X.; Smyth, F.; Papas, T.S.; Watson, D.K.; Schultz, R.A.

Db 291 AMPHKKSGTFKQYVRRAELNKKDPYIPAAALAGTSGSPIDQMQLLELTDKSCQSF 350
QY 329 ITWEGTGERKMDPDDEVARWGERKSKPMNNTDKLSRALRYYYDKNIMTKVHPRESSMY 388
Db 351 ISMTGDMEERKLPDDEVARWGERKSKPMNNTDKLSRALRYYYDKNIMTKVHPRESSMY 402
QY 389 KYPSPDLPMYSSHYCKRYAKKF 409
Db 403 -----AGKRYVRR 411

RESULT 14

A48146
GA-binding protein alpha chain - human
N/Alternate names: GABPA; nuclear respiratory factor-2 alpha chain; transcription factor
C/Species: Homo sapiens (man)
C/Date: 21-Jan-1994 #sequence, revision 23-May-1997 #text_change 16-Jul-1999
C/Accession: A48146; A46303; A55903
R/Matlab: H.; Sawada, J.; Yano, K.; Yamaguchi, K.; Goto, M.; Handa, H.
Mol. Cell. Biol. 13, 1385-1391, 1993
A/Title: cDNA cloning of transcription factor E4F1 subunits with Ets and notch motifs.
A/Reference number: A48146; MUID:93180783; PMID:8441384
A/Accession: A48146
A/Status: not compared with conceptual translation
A/Molecule type: mRNA; protein
A/Residues: 1-454 <MAT>
A/Note: sequence extracted from NCBI backbone (NCBI:125762)
A/Note: parts of this sequence were determined by protein sequencing
R/Virbasius, J.V.; Virbasius, C.A.; Scarpulla, R.C.
Genes Dev. 7, 380-392, 1993
A/Title: Identity of GABP with NRE-2, a multisubunit activator of cytochrome oxidase exp
A/Reference number: A46303; MUID:93194058; PMID:8383622
A/Accession: A46303
A/Molecule type: protein
A/Residues: 237-248; 429-448 <VIR>
R/Gardot, P.; Pastier, D.; Lacorte, J.M.; Mangeney, M.; Zanits, V.I.; Chambaz, J.
Biochemistry 33, 12139-12148, 1994
A/Title: Purification and characterization of nuclear factors binding to the negative re
ets-related protein.
A/Reference number: A55903; MUID:95001929; PMID:7918435
A/Accession: A55903
A/Molecule type: protein
A/Residues: 26-34, 'A', '36', 'X', '38', 'X', '40-46, 151-165, 167-174; 350-359 <CAR>
C/Genetics:
A/Gene: GDB:GABPA; E4F1-60; E4F1A
A/Cross-references: GDB:138476; OMIM:600609
A/Map position: 21q21-21q22.1
C/Complex: GA-binding protein is a heterotetramer of two alpha and two beta-type chains.
C/Function:
A/Description: a transcription factor that binds (via the alpha chain) to GA-rich promot
A/Pathway: known to promote transcription of apolipoprotein A-II, cytochrom c oxidase c
C/Superfamily: transcription factor elg; ets DNA-binding domain homology; ets RII regu
C/Keywords: DNA binding; nucleus; transcription factor
F/174-245/Domain: ets RII regulatory region homology <ETR>
F/322-400/Domain: ets DNA-binding domain homology <ETS>

Query Match 17.1%; Score 443.5; DB 1; Length 454;
Best Local Similarity 33.0%; Pred. No. 1, 1e-23;
Matches 112; Conservative 48; Mismatches 94; Indels 85; Gaps 13;

QY 111 KHPPPMNTNER-----RVYPADPTLMSSTDHVRQWLEMAVKEYGLPD 154
Db 150 KHITTSDETSQVTRMAALEGRRKQEDRIKIPYDIOWSTQVLMHVVWVWMEFSMTD 209
QY 155 VDLIFONIGKELCKMTKDFORLPPSYNADILSHLHREPLPHLNSD-----DVD 209
Db 210 IDLTTLNIGRELCSLNQEDFQRVP--RGEILMSHLELRKVLASQEQKMEIYITID 266
QY 210 KALONSRLMHAARNTGATFTFNTSYPRATORITTRPDLPEQARRSAMTSHSPTQS 269
Db 267 QPVO-----IIPASVGSAT-----PTTI 284
QY 270 KATQPSSTYPTKEDQRPOLDPIQILGPTSSRLANPSGQIOLMQLLELSDSSNSNCI 329

Db 285 KVIN-SSAKAKAVQ--RAPRISGEDRSSP--GNRTGN--NGQIOLMQLLELTDKARDCI 339
QY 330 TWEGTNEEFKMTDDEVARWGERKSKPMNNTDKLSRALRYYYDKNIMTKVHPRESSMYK 389
Db 340 SWMGDEGEFKLNQPELVAKQKGRKKNPTMYEKLRSALRYYYGDMICKVQ--GRRYVK 398
QY 390 YPSDLPYMSSTYH-----KRYAKFDFHIGIAQ 416
Db 399 FVCDLKTILIGYSAELNRLVTECBQKILA--KMQHLHIGIAQ 436

RESULT 15

A40858
GA-binding protein alpha chain - mouse
N/Alternate names: GABPA; nuclear respiratory factor-2 alpha chain; transcription fac
C/Species: Mus musculus (house mouse)
C/Date: 28-Feb-1992 #sequence, revision 23-May-1997 #text_change 16-Jul-1999
C/Accession: A40858
R/Lamarca, K.; Thompson, C.C.; Byers, B.P.; Walton, E.M.; McKnight, S.L.
Science 253, 789-792, 1991
A/Title: Identification of ets- and notch-related subunits in GA binding protein.
A/Reference number: A40858; MUID:91343912; PMID:1876836
A/Accession: A40858
A/Molecule type: mRNA; protein
A/Residues: 1-454 <LAM>
A/Cross-references: GB:W4515; NID:9193382; PIDN:AA53030.1; PID:9193383
A/Note: parts of this sequence were determined by protein sequencing
C/Genetics:
A/Gene: GABPA
A/Map position: 16 47.0
C/Complex: GA-binding protein is a heterotetramer of two alpha and two beta-type cha
C/Function:
A/Description: a transcription factor that binds (via the alpha chain) to GA-rich pro
A/Pathway: known to promote transcription of apolipoprotein A-II, cytochrom c oxidase
C/Superfamily: transcription factor elg; ets DNA-binding domain homology; ets RII reg
C/Keywords: DNA binding; nucleus; transcription factor
F/174-245/Domain: ets RII regulatory region homology <ETR>
F/322-400/Domain: ets DNA-binding domain homology <ETS>

Query Match 17.1%; Score 441.5; DB 1; Length 454;
Best Local Similarity 28.6%; Pred. No. 1, 5e-23;
Matches 126; Conservative 66; Mismatches 127; Indels 119; Gaps 17;

QY 15 DQSLFECAVSGPHLAKTEMTASSSEYQOTSKMSRPYPODMLSQPARVTIKMECPNQ 74
Db 78 DRSLEFDG-----VKIDGTYQVLSQVYISQGMERKL--NILEIVTAIEVEVDPD- 127
QY 75 VNGSRNSPDDCSVAKGKMWSSSDNMGNTGSIWEKHIPPMTNER-----123
Db 128 ---AHNAEAAHVAEEAQTITLDGT-----KHITTSDETSQVTRMAALEGY 173
QY 124 ---RVYPADPTLMSSTDHVRQWLEMAVKEYGLPDVILFQINDKELCKMTKDFOR 178
Db 174 RKQERIGIPDPIRMSTQVLMHVVWVWMEFSMTDIDLTTLNIGRELCSLNQEDFQ 232
QY 179 LTPSYNADILSHLHREPLPHLNSD-----DVDKALONSRLMHAARNTGATFTFPN 233
Db 233 RVP--RGEILMSHLELRKVLASQEQKMEIYITDPVOIIP-----273
QY 234 TSYVPEATQRTTRPDLPEQARRSAMTSHSPTQSKATQPSSTYPTKEDQRPOLDPYQ 293
Db 274 -ASVPAPAT-----PTTIKVINSSAKA-----KVQSRP-R 301
QY 294 ILGPTSSRLANPG-----SGQIOLMQLLELSDSSNSNCITWEGTGERKMDPDDEVAR 348
Db 302 ISGEDRS--SPGKRTGNNGQIOLMQLLELTDKARDCISWGDGEERKLNPDLVAQ 358
QY 349 RMGERKSKPMNNTDKLSRALRYYYDKNIMTKVHPRESSMYKYPSPDLPMYSSHYG-----402
Db 359 KMGQRKKNPTMYEKLRSALRYYYGDMICKVQ--GRFYVKFVCDLKTILIGYSAELNRL 417
QY 403 -----KRYAKFDFHIGIAQ 416

Db 418 VIECEQKILA-RMQLHGIAQ 436

Search completed: November 9, 2002, 16:54:39
Job time : 32.3864 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 9, 2002, 15:37:21 ; Search time 21.0958 Seconds
(without alignments)
939.793 Million cell updates/sec

Title: US-09-902-772-4

Perfect score: 2588
Sequence: 1 MASTKEALSVSEDSLFE.....IYPNRLPAAHMPSHLCTYY 478

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	2436	94.1	478 1	ERG_CHICK
2	2180.5	84.3	462 1	ERG_HUMAN
3	1482	57.3	452 1	FLI1_HUMAN
4	1474	57.0	452 1	FLI1_MOUSE
5	1440.5	55.7	453 1	FLI1_XENLA
6	1253.5	48.4	272 1	ERG_MOUSE
7	596	23.0	173 1	ERG_LYTV
8	551.5	21.3	475 1	ETS6_DROME
9	502.5	19.4	184 1	ETS3_DROME
10	455	17.6	464 1	ELG_DROME
11	444	17.2	438 1	ETLA_XENLA
12	443.5	17.1	454 1	GABA_HUMAN
13	441.5	17.0	454 1	GABA_MOUSE
14	441	17.0	479 1	ETS7_CHICK
15	440.5	17.0	441 1	ETS1_HUMAN
16	435	16.8	472 1	ETS2_XENLA
17	434.5	16.8	485 1	ETS8_CHICK
18	433.5	16.8	441 1	ETS1_RAT
19	433.5	16.8	441 1	ETS3_CHICK
20	432.5	16.7	469 1	ETS2_HUMAN
21	430.5	16.6	440 1	ETS1_MOUSE
22	427	16.5	468 1	ETS2_MOUSE
23	426	16.5	472 1	ETS2_XENLA
24	425.5	16.4	669 1	MYBE_AVILE
25	351.5	13.6	718 1	PNT2_DROME
26	326	12.6	268 1	ETV1_XENLA
27	317	12.2	452 1	ETV1_HUMAN
28	314.5	12.2	250 1	ETV3_HUMAN
29	313.5	12.1	732 1	POK_DROME
30	303.5	11.7	548 1	ERF_HUMAN
31	303.5	11.7	551 1	ERF_MOUSE
32	302	11.7	623 1	PNT1_DROME
33	300.5	11.6	110 1	ETS2_LYTV

34	294.5	11.4	485 1	ETV6_MOUSE	P97360 mus musculu
35	292.5	11.3	477 1	ETV1_HUMAN	P50549 homo sapien
36	288	11.1	477 1	ETV1_MOUSE	P41164 mus musculu
37	280.5	10.8	510 1	ERM_HUMAN	P41161 homo sapien
38	278.5	10.8	335 1	ETV2_MOUSE	P41163 mus musculu
39	277.5	10.7	341 1	ETV2_HUMAN	C00321 homo sapien
40	273	10.5	341 1	TEL2_HUMAN	Q9Y603 homo sapien
41	266	10.3	355 1	ETV4_MOUSE	P28322 mus musculu
42	264	10.2	430 1	ELK4_MOUSE	P41168 mus musculu
43	263.5	10.2	551 1	ETV4_HUMAN	P43268 homo sapien
44	261.5	10.1	428 1	ELK1_HUMAN	P19419 homo sapien
45	258.5	10.0	429 1	ELK1_MOUSE	P41969 mus musculu

ALIGNMENTS

RESULT 1
ERG_CHICK ID ERG_CHICK STANDARD; PRT; 478 AA.
AC Q90837;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Transcriptional regulator Erg.
GN ERG.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Spleen;
RX MEDLINE=95329425; Pubmed=7605748;
RA Duterque-Cogullaud M.;
RA Mesodermal expression of the chicken erg gene associated with
RT precartilaginous condensation and cartilage differentiation.;
RL Mech. Dev. 50:17-28(1995).
CC - FUNCTION: ACTS AS A TRANSCRIPTIONAL ACTIVATOR.
CC - SUBCELLULAR LOCATION: Nuclear.
CC - TISSUE SPECIFICITY: EXPRESSED IN MESODERM- AND, TO A LESSER
CC EXTENT, IN ECTODERM-DERIVED TISSUES.
CC - SIMILARITY: BELONGS TO THE ETS FAMILY.
CC - SIMILARITY: CONTAINS 1 POINTED (PNT) DOMAIN.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X77159; CAA54404.1; -.
CC HSSP: O01543; IFLI.
CC InterPro: IPR000418; ETS.
CC InterPro: IPR002341; HSF.ETS.
CC InterPro: IPR003118; SAM_PNT.
CC Pfam: PF00178; Ets; 1.
CC Pfam: PF02198; SAM_PNT; 1.
CC PRINTS: PR00454; ETSDOMAIN.
CC SMART: SM00413; ETS; 1.
CC SMART: SM00251; SAM_PNT; 1.
CC PROSITE: PS00345; ETS_DOMAIN_1; 1.
CC PROSITE: PS00346; ETS_DOMAIN_2; 1.
CC PROSITE: PS50061; ETS_DOMAIN_3; 1.
CC Transcription regulation: Activator; Nuclear protein; DNA-binding.
FT DOMAIN 115 199 POINTED.
FT DNA_BIND 310 390 ETS-DOMAIN.
SQ SEQUENCE 478 AA; 53913 MW; 8DFDB243EDB623A7 CRC64;

Query Match 94.1%; Score 2436; DB 1; Length 478;
 Best Local Similarity 92.0%; Pred. No. 4,4e-163;
 Matches 458; Conservative 0; Mismatches 0; Indels 40; Gaps 2;

QY 1 MASTIKREALSVSEDOSELECAVSPHLAKTEMTASSSEYGGTSMKSPRVPOODMLSDP 60
 DB 1 MASTIKREALSVSEDOSELECAVSPHLAKTEMTASSSEYGGTSMKSPRVPOODMLSDP 60
 QY 61 PARVITKMECNPNQVNGSRNSPDDCSVAKGKVVSSSDVNGVNGYGYMEKHIPPNNMTT 120
 DB 61 PARVITKMECNPNQVNGSRNSPDDCSVAKGKVVSSSDVNGVNGYGYMEKHIPPNNMTT 120
 QY 121 NERRVIVPADPTLMSVDHVRQWLEMAVKEYGLPDVDILFQNDGKELCKMTKDDPQRLTP 180
 DB 121 NERRVIVPADPTLMSVDHVRQWLEMAVKEYGLPDVDILFQNDGKELCKMTKDDPQRLTP 180
 QY 181 PSYNADILSHLYRETPPLPHLTSDDVDKALQNSRLHARNTGGATFIFPNTSYYPEA 240
 DB 181 PSYNADILSHLYRETPPLPHLTSDDVDKALQNSRLHARNTGGATFIFPNTSYYPEA 240
 QY 241 TORITTRPDLPEQARASAMTSHSHPTQSKATQPSSTVPTKEDQRPOLDPYOILGPTSS 300
 DB 241 TORITTRPDLPEQARASAMTSHSHPTQSKATQPSSTVPTKEDQRPOLDPYOILGPTSS 300
 QY 301 RIANPSSGOIQIOMQFLLELSDSSNCTTWSTNGEFKMTDPDEVARWGERKSKPPNN 360
 DB 301 RIANPSSGOIQIOMQFLLELSDSSNCTTWSTNGEFKMTDPDEVARWGERKSKPPNN 360
 QY 361 YKLSRALRYRYDKNIMTY-----HPPSSMYTYPSPDLFYNSSX 400
 DB 361 YKLSRALRYRYDKNIMTYHGKRYAKKDFHGIAQALOPHPRESSMYTYPSPDLFYNSSX 420
 QY 401 HGKRYAKKDFHGIAQALOPHPAPKMFVAPHPALPTVSSSFFAAPNPYNSPTGTY 460
 DB 401 -----HAHPQKMFVAPHPALPTVSSSFFAAPNPYNSPTGTY 460
 QY 461 PMTRLPAAHMPSHLGY 478
 DB 461 PMTRLPAAHMPSHLGY 478

RESULT 2
 ERG_HUMAN STANDARD; PRT; 462 AA.
 AC P1308;
 DC 01-JUL-1989 (Rel. 11, Created)
 DT 01-JUL-1989 (Rel. 11, Last sequence update)
 DR 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Transcriptional regulator ERG (Transforming protein ERG).
 GN ERG.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM ERG-2).
 RX MEDLINE=87263429; PubMed=3299708;
 RA Rao V.N., Papas T.S., Shyam E., Reddy P.;
 RT "erg, a human ets-related gene on chromosome 21: alternative splicing,
 RT polyadenylation, and translation.";
 RL Science 237:635-639(1987).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM ERG-1).
 RX MEDLINE=87317608; PubMed=3476934;
 RA Reddy E.S.P., Rao V.N., Papas T.S.;
 RT "The erg gene: a human gene related to the ets oncogene.";
 RL Proc. Natl. Acad. Sci. U.S.A. 84:6131-6135(1987).
 RN [3]
 RP CHROMOSOMAL TRANSLOCATION.
 RX MEDLINE=94356859; PubMed=8076344;
 RA Dunn T., Pralissman L., Hageg N., Viola M.V.;
 RT "ERG gene is translocated in an Ewing's sarcoma cell line.";
 RL Cancer Genet. Cytogenet. 76:19-22(1994).

RN [4]
 RP CHROMOSOMAL TRANSLOCATION.
 RX MEDLINE=94243799; PubMed=8187069;
 RA Ichikawa H., Shimizu K., Hayashi Y., Ohki M.;
 RT "An RNA-binding protein gene, Tls/Fus, is fused to ERG in human
 RT myeloid leukemia with t(16;21) chromosomal translocation.";
 RL Cancer Res. 54:2865-2868(1994).
 CC -1- FUNCTION: ACTS AS A TRANSCRIPTIONAL ACTIVATOR (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: ERG-1 AND ERG-2 (SHOWN HERE);
 CC ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- DISEASE: A FORM OF ACUTE MYELOID LEUKEMIA (AML) IS CHARACTERIZED
 CC BY A CHROMOSOMAL TRANSLOCATION T(16;21)(P11;Q22) WHICH INVOLVES
 CC ERG AND FUS.
 CC -1- SIMILARITY: BELONGS TO THE ETS FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 POINTED (PNT) DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; M17254; AAA52398.1; -;
 DR EMBL; M21535; AAA35811.1; -;
 DR PIR; A29515; TVHUEG.
 DR HSSP; C01543; 1FLI.
 DR TRANSFAC; T00265; -;
 DR TRANSFAC; T02129; -;
 DR TRANSFAC; T02130; -;
 DR GeneW; HGNC:3446; ERG.
 DR MIM; 165080; -;
 DR InterPro; IPR000418; Ets.
 DR InterPro; IPR002341; HSF_ETs.
 DR InterPro; IPR003118; SAM_PNT.
 DR Pfam; PF00178; Ets; 1.
 DR Pfam; PF02198; SAM_PNT; 1.
 DR PRINTS; PR00454; ETSDOMAIN.
 DR SMART; SM00251; SAM_PNT; 1.
 DR PROSITE; PS00345; ETS_DOMAIN_1; 1.
 DR PROSITE; PS00346; ETS_DOMAIN_2; 1.
 DR PROSITE; PS00611; ETS_DOMAIN_3; 1.
 KW Transcription regulation; Activator; Nuclear protein; DNA-binding;
 KW Alternative splicing; Proto-oncogene; Chromosomal translocation.
 FT DOMAIN 122 206 POINTED.
 FT DNA_BIND 294 374 ETS-DOMAIN.
 FT VARSPIC 1 99 MISSING (IN ISOFORM ERG-1).
 SQ SEQUENCE 462 AA; 52031 MW; B29F14B0F5C2C697 CRC64;

Query Match 84.3%; Score 2180.5; DB 1; Length 462;
 Best Local Similarity 82.9%; Pred. No. 3e-145;
 Matches 413; Conservative 9; Mismatches 11; Indels 65; Gaps 4;

QY 2 ASTIKREALSVSEDOSELECAVSPHLAKTEMTASSSEYGGTSMKSPRVPOODMLSDP 61
 DB 9 AAHIKREALSVSEDOSELECAVSPHLAKTEMTASSSEYGGTSMKSPRVPOODMLSDP 68
 QY 62 PARVITKMECNPNQVNGSRNSPDDCSVAKGKVVSSSDVNGVNGYGYMEKHIPPNNMTT 121
 DB 69 PARVITKMECNPNQVNGSRNSPDDCSVAKGKVVSSSDVNGVNGYGYMEKHIPPNNMTT 128
 QY 122 ERRVIVPADPTLMSVDHVRQWLEMAVKEYGLPDVDILFQNDGKELCKMTKDDPQRLTP 181
 DB 129 ERRVIVPADPTLMSVDHVRQWLEMAVKEYGLPDVDILFQNDGKELCKMTKDDPQRLTP 188
 QY 182 SYNADILSHLYRETPPLPHLTSDDVDKALQNSRLHARNTGGATFIFPNTSYYPEAT 241
 DB 189 SYNADILSHLYRETPPLPHLTSDDVDKALQNSRLHARNTGGATFIFPNTSYYPEAT 231

QY 242 QRTTRPDLPEQARRSAMTSHIPT-QSKATQSSSTVPTEDQRPOLDPQILGPTSS 300
 Db 232 -----DLPEPPRRSRAMTHGHPTPOSKAAPSPSTVPTEDQRPOLDPQILGPTSS 284
 QY 301 RLAMPSSGOIOLMOFLLELSDSSNSNCITWEGTNGEFGKMTDPPEVARRMERSKPMN 360
 Db 285 RLAMPSSGOIOLMOFLLELSDSSNSNCITWEGTNGEFGKMTDPPEVARRMERSKPMN 344
 QY 361 YDKLSRALRYYYDKNIMTKVHPPESSMKYPPSCLPYMSSYKGRYAKFDFHGIAQALQP 420
 Db 345 YDKLSRALRYYYDKNIMTKV-----HGKRYAYKDFHGIAQALQP 384
 QY 421 -----HAHQKMFVAPHPALPYTSSSFPAAPNPTWNSPTGCIY 460
 Db 385 HPPESSLYKYPSPCLPYMSSYKGRYAKFDFHGIAQALQP 444
 QY 461 PNTRLPAHMPSHLGTYX 478
 Db 445 PNTRLPSTHMPSHLGTYX 462

RESULT 3

FL1L_HUMAN STANDARD; PRT; 452 AA.
 ID FL1L_HUMAN STANDARD; PRT; 452 AA.
 AC 001543; Q14319; Q9UB07; Q92480;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Friend leukemia integration 1 transcription factor (Flt-1 proto-oncogene) (ERGB transcription factor).
 GN FL1L.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE=Bone marrow;
 RX MEDLINE=92396239; PubMed=1522903;
 RA Delattre O., Zucman J., Plougastel B., Desmaza C., Melot T., Peter M., Kovar H., Joubert I., de Jong P., Rouleau G.;
 RT "Gene fusion with an ETS DNA-binding domain caused by chromosome translocation in human tumours.";
 RL Nature 359:162-165(1992).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RX MEDLINE=93075640; PubMed=1445800;
 RA Watson D.K., Smyth F.E., Thompson D.M., Cheng J.Q., Testa J.R., Pappas T.S., Seth A.;
 RT "The ERGB/Flt-1 gene: isolation and characterization of a new member of the family of human ETS transcription factors.";
 RL Cell Growth Differ. 3:705-713(1992).
 RN [3]
 RP REVISIONS.
 RA Watson D.;
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RX MEDLINE=93007976; PubMed=1394211;
 RA Prasad D.D., Rao V.N., Reddy E.S.;
 RT "Structure and expression of human Flt-1 gene.";
 RL Cancer Res. 52:5833-5837(1992).
 RN [5]
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
 RC TISSUE=Blood;
 RX MEDLINE=93176799; PubMed=8439553;
 RA Thomas R., May W., Denny C., Raabind W., Moore J., Maki R.A., Beck E., Klemsz W.J.;
 RT "Human Flt-1 localizes to chromosome 11q24 and has an aberrant transcript in neuroepithelioma.";
 RL Biochim. Biophys. Acta 1172:155-158(1993).
 RN [6]
 RP SEQUENCE FROM N.A. (ISOFORM 1).

RA Ubhl B.T.S., Rainey D.R., Meredith D.M.;
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 RN [7]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE=Lymph;
 RA Strausberg R.;
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 RN [8]
 RP SEQUENCE OF 58-452 FROM N.A.
 RX MEDLINE=98426231; PubMed=9751743;
 RA Zucman-Rossi J., Legoux P., Victor J.M., Lopez B., Thomas G.;
 RT "Chromosome translocation based on illegitimate recombination in human tumors.";
 RL Proc. Natl. Acad. Sci. U.S.A. 95:11786-11791(1998).
 RN [9]
 RP SEQUENCE OF 278-301 FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=95352541; PubMed=7542907;
 RA Bhagirath T., Abe S., Nojima T., Yoshida M.C.;
 RT "Molecular analysis of a t(11;22) translocation junction in a case of Ewing's sarcoma.";
 RL Genes Chromosomes Cancer 13:126-132(1995).
 RN [10]
 RP STRUCTURE BY NMR OF 276-373.
 RX MEDLINE=95292091; PubMed=773776;
 RA Liang H., Mao X., Olejniczak E.T., Nettesheim D.G., Yu L., Meadows R.P., Thompson C.B., Resik S.W.;
 RT "Solution structure of the ets domain of Flt-1 when bound to DNA.";
 RL Nat. Struct. Biol. 1:871-875(1994).
 CC -1- FUNCTION: SEQUENCE SPECIFIC TRANSCRIPTIONAL ACTIVATOR. RECOGNIZES THE DNA SEQUENCE 5'C[CA]GGAGT-3'.
 CC -1- SUBUNIT: CAN FORM HOMODIMERS OR HETERODIMERS WITH ETV6/TEL.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: 1 (SHOWN HERE) AND 2; SEEM TO BE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- DISEASE: A FORM OF EWING'S SARCOMA IS CHARACTERIZED BY A CHROMOSOMAL TRANSLOCATION T(11;22)(Q24,Q12) WHICH INVOLVES FL1L AND EWS.
 CC -1- SIMILARITY: BELONGS TO THE ETS FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 POINTED (PNT) DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X67001; CAA47399.1; -;
 DR EMBL: M98833; AAA35812.2; -;
 DR EMBL: S45205; AAB23637.1; -;
 DR EMBL: M93255; AAA58479.1; -;
 DR EMBL: M93255; AAA58480.1; -;
 DR EMBL: AY029368; AAK50443.1; -;
 DR EMBL: BC010115; AAH01670.1; -;
 DR EMBL: BC010115; AAH01115.1; -;
 DR EMBL: Y17293; CAA76731.1; -;
 DR EMBL: D38408; BAA07463.1; ALT_TERM.
 DR PDB: 1FLI; 15-SEP-95.
 DR TRANSFAC: T02066; -;
 DR Genew: HGNC:3749; FL1L.
 DR MIM: 193067; -;
 DR Interpro: IPR000418; ETS.
 DR Interpro: IPR002341; HSF_ET5.
 DR Interpro: IPR003118; SAM_PNT.
 DR Pfam: PF00178; ETS_1.
 DR Pfam: PF02198; SAM_PNT_1.
 DR PRINTS: PR00413; ETS_1.
 DR SMART: SM00413; ETS_1.
 DR SMART: SM00251; SAM_PNT_1.
 DR PROSITE: PS00345; ETS_DOMAIN_1; 1.
 DR PROSITE: PS00346; ETS_DOMAIN_2; 1.
 DR PROSITE: PS00346; ETS_DOMAIN_2; 1.


```

QY 119 TTNRRRVTPADPTLMSTDHVRQMLEAVKEXGLPDVUILLFONTDKELCKMTKDFOR 178
D 118 TTNRRRVTPADPTLMSTDHVRQMLEAVKEXGLPDVUILLFONTDKELCKMTKDFOR 177
QY 179 LTPSYNADILSHLHYRETPPLPHLTSDVDVKALQNSPRLMHANTGATFFPNTSYR 238
D 178 ATSAVNEVLSHLSTYRESL-----LANTYSHT 208
QY 239 EATORTTTPDLPYEQARRSAMTSHTPTOSKA-TOPSSSTVPKTEDORPOLDPYQILGP 297
D 209 DQSSRLNFKEDPSYDSVARGAMNNMNGKLSPLLGSSQTMKNTBGRQPDPPYQILGP 268
QY 298 TSSRLANPGSGQIQLMQFLLELSDSSNSCITWEGTNGEKKMTDPDEVARRWGERKSKP 357
D 269 TSSRLANPGSGQIQLMQFLLELSDSSNSCITWEGTNGEKKMTDPDEVARRWGERKSKP 328
QY 358 NMNVDKLSRALRYRDKIMTKVHPRESSMKYPSDLPYMSYSGKRYAYKEDFHGIAQ 417
D 329 NMNVDKLSRALRYRDKIMTKV-----HGKRYAYKEDFHGIAQ 368
QY 418 LQP-----HAHPQKNFVAPHPALPVTSSEFAAPNPYNSPTG 457
D 369 LQPHPTETSMYKYPDSISYMPSYAHQKYNFVSHPSMVPYSSSEFGAASQYWTSP 428
QY 458 GYIPN---TRLPAAMPSHLCTTY 478
D 429 GYIPNPSVPRHPNTHVPSHLGSY 452

```

RESULT 5

```

FLII_XENLA STANDARD; PRT; 453 AA.
ID FLII_XENLA
AC P4157;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Retroviral integration site protein FLI-1 homolog.
GN FLI.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94206844; PubMed=8155576;
RA Meyer D., Wolff C.M., Stiegler P., Senan F., Befort N.,
RA Befort J.J., Remy P.;
RT "Xl-fl, the Xenopus homologue of the fl-1 gene, is expressed during
RT embryogenesis in a restricted pattern evocative of neural crest cell
RT distribution."
RL Mech. Dev. 44:109-121(1993).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED DURING EMBRYOGENESIS.
CC -1- SIMILARITY: BELONGS TO THE ETS FAMILY.
CC -1- SIMILARITY: CONTAINS 1 POINTED (PNT) DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: X66979; GAA47389.1; -
DR HSP: Q01543; 1FLI.
DR TRANSFAC: T02067; -
DR InterPro: IPR000418; Ets.
DR InterPro: IPR002341; HSF_ET.
DR InterPro: IPR003118; SAM_PNT.
DR Pfam: PF00178; Ets; 1.
DR Pfam: PF02198; SAM_PNT; 1.

```

```

DR PRINTS: PR00454; ETSDOMAIN.
DR SMART: SM00413; ETS; 1.
DR SMART: SM00251; SAM_PNT; 1.
DR PROSITE: PS00345; ETS_DOMAIN_1; 1.
DR PROSITE: PS00346; ETS_DOMAIN_2; 1.
DR PROSITE: PS50061; ETS_DOMAIN_3; 1.
DR Transcription regulation; Activator; DNA-binding; Nuclear protein.
FW DOMAIN 113 197 POINTED.
FT DNA_BIND 282 362 ETS-DOMAIN.
SQ SEQUENCE 453 AA; 51015 MM; A6E6FC6CE42D4A CRC64;
Query Match 55.7%; Score 1440.5; DB 1; Length 453;
Best Local Similarity 57.4%; Pred. No. 1.5e-93;
Matches 290; Conservative 50; Mismatches 86; Indels 79; Gaps 11;

```

```

QY 1 MASTIKELATVSEDSQTEFCAYG-SPLIAKTEMASSSEFGOTSKSPRPOODWISO 59
D 1 MDGITEALSYVSDQSLFDSATGASSHLSKADMTASNPDPGPKINPIPPQDDWINO 60
QY 60 PPARVTIMECNPNQVNGSRNSPDDCSYAKGKKNVSSSDNYGMNGSYMEERH-IPPPNM 118
D 61 -PMRYNIKREY--EHMNGSRSPVDCSINKCSKLIGSEGNAMTY-TYMDKNGPPPPNM 116
QY 119 TTNRRRVTPADPTLMSTDHVRQMLEAVKEXGLPDVUILLFONTDKELCKMTKDFOR 178
D 117 TTNRRRVTPADPTLMSTDHVRQMLEAVKEXGLPDVUILLFONTDKELCKMTKDFOR 176
QY 179 LTPSYNADILSHLHYRETPPLPHLTSDVDVKALQNSPRLMHANTGATFFPNTSYR 238
D 177 STSIYNTVEVLSHLNLYLDS-----SSSIGVTYQAH 208
QY 239 EATORTTTPDLPYEQARRSAM-TSHSPTOSKATOPSSSTVPKTED-ORPOLDPYQILG 296
D 209 DQSSRLNFKEDPSYDSVARGAMNNMNGKLSPLLGSSQTMKNTBGRQPDPPYQILG 268
QY 297 TSSRLANPGSGQIQLMQFLLELSDSSNSCITWEGTNGEKKMTDPDEVARRWGERKSK 356
D 269 TSSRLANPGSGQIQLMQFLLELSDSSNSCITWEGTNGEKKMTDPDEVARRWGERKSK 328
QY 357 NMNVDKLSRALRYRDKIMTKVHPRESSMKYPSDLPYMSYSGKRYAYKEDFHGIAQ 416
D 329 NMNVDKLSRALRYRDKIMTKV-----HGKRYAYKEDFHGIAQ 368
QY 417 ALQP-----HAHPQKNFVAPHPALPVTSSEFAAPNPYNSPT 456
D 369 ALQPHPTETSMYKYPSESYMPSYAHQKYNFVSHPSMVPYSSSEFGAASQYWTSP 428
QY 457 GYIPN---TRLPAAMPSHLCTTY 478
D 429 ANIYPNPVPRHPNTHVPSHLGSY 453

```

RESULT 6

```

ERG_MOUSE STANDARD; PRT; 272 AA.
ID ERG_MOUSE
AC P81270;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Transcriptional regulator ERG (Fragment).
GN ERG OR ERG-3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94019387; PubMed=8413305;
RA Rivaiera R.R., Stulver M.H., Steenbergen R., Murie C.;
RT "Ets proteins: new factors that regulate immunoglobulin heavy-chain
RT gene expression."
RL Mol. Cell. Biol. 13:7163-7169(1993).
CC -1- FUNCTION: ACTS AS A TRANSCRIPTIONAL ACTIVATOR.

```

```

CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE ETS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: S66169; AAB28525.1; -
DR HSSP: Q01543; 1FLL.
DR MGD: MGI:95415; ETS.
DR InterPro: IPR000418; Ets.
DR InterPro: IPR002341; HSF_ETs.
DR Pfam: PF00178; Ets; 1.
DR SMART: SM00413; ETS; 1.
DR PROSITE: PS00345; ETS_DOMAIN_1; 1.
DR PROSITE: PS00346; ETS_DOMAIN_2; 1.
DR PROSITE: PS50061; ETS_DOMAIN_3; 1.
DR PROSITE: PS50061; ETS_DOMAIN_3; 1.
KW Transcription regulation; Activator; Nuclear protein; DNA-binding.
FT NON_TER 1 1
FT DNA_BIND 138 218 ETS-DOMAIN.
FT NON_TER 272 272
SQ SEQUENCE 272 AA; 30907 MW; 895A512C1B14B4A8 CRC64;

Query Match 48.4%; Score 1253.5; DB 1; Length 272;
Best local similarity 82.9%; Pred. No. 9.2e-81;
Matches 242; Conservative 1; Mismatches 8; Indels 41; Gaps 3;

QY 174 DDQRRLPSYNADILLSHLYLREPLRLTSDVDKALONSRLMAHARTGCATLFPN 233
DB 1 DDQRRLPSYNADILLSHLYLREPLRLTSDVDKALONSRLMAHARTGCATLFPN 60

QY 234 TSYPPATQRTTRPDLPEYQARSAMTSSHPT-OSKATOPSSSTVYPKREDROPOLDPY 292
DB 61 TSYPPATQRTTRPDLPEYPPRSAMTGHSLTPQSKAQSPSAVPRKEDROPOLDPY 120

QY 293 QILGPTSSRLANFGSGQIOLMOLFLELLSDSSNSNCITWEGTNGEERKMTDPDEVARRWGE 352
DB 121 QILGPTSSRLANFGSGQIOLMOLFLELLSDSSNSNCITWEGTNGEERKMTDPDEVARRWGE 180

QY 353 RSKSPMANYDKLSRALRYVYDKRMKRVHPRESSMYKYSDDLPMSSYHGKRAKYKFDH 412
DB 181 RSKSPMANYDKLSRALRYVYDKRMKRVHPRESSMYKYSDDLPMSSYHGKRAKYKFDH 220

QY 413 GIAQALQP-----HAHPQKMFVAHPHPALPVTSSSF 444
DB 221 GIAQALQPHPPRESSLYKYPDDLPMGSIYAHHPQKMFVAHPHPALPVTSSSF 272

RESULT 7
ERG_LYTVA STANDARD: PRT: 173 AA.
AC 001414:
DT 01-FEB-1995 (rel. 31, Created)
DT 01-FEB-1995 (rel. 31, Last sequence update)
DT 15-DEC-1998 (rel. 37, Last annotation update)
DE Transcriptional regulator ERG homolog (Fragment).
GN ERG.
OS Eukaryotus variegatus (Sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidae; Echininoidea; Echinacea; Temnopileuroidea; Toxopneustidae;
OC Lytechinus.
OX NCBI_TaxID=7654;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE-93091246; PubMed-1457815;
RA Q1 S., Chen Z.O., Papas T.S., Laubenberger J.A.;
RA "The sea urchin erg homolog defines a highly conserved erg-specific
RA domain."
RA DNA Seq. 3:127-129(1992).

```

```

CC -1- FUNCTION: ACTS AS A TRANSCRIPTIONAL ACTIVATOR.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE ETS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M81067; AAA68905.1; -
DR HSSP: Q01543; 1FLL.
DR InterPro: IPR000418; Ets.
DR InterPro: IPR002341; HSF_ETs.
DR Pfam: PF00178; Ets; 1.
DR SMART: SM00413; ETS; 1.
DR PROSITE: PS00345; ETS_DOMAIN_1; 1.
DR PROSITE: PS00346; ETS_DOMAIN_2; 1.
DR PROSITE: PS50061; ETS_DOMAIN_3; 1.
DR PROSITE: PS50061; ETS_DOMAIN_3; 1.
KW Transcription regulation; Activator; Nuclear protein; DNA-binding.
FT NON_TER 1 1
FT DNA_BIND 1 84 ETS-DOMAIN.
FT NON_TER 84 84
SQ SEQUENCE 173 AA; 19690 MW; F85D496DD58ABF3D CRC64;

Query Match 23.0%; Score 596; DB 1; Length 173;
Best local similarity 61.9%; Pred. No. 4.5e-35;
Matches 120; Conservative 13; Mismatches 17; Indels 44; Gaps 5;

QY 307 SGQIOLMOLFLELLSDSSNSNCITWEGTNGEERKMTDPDEVARRWGERKSPNNYDKLSR 366
DB 1 SGQIOLMOLFLELLSDSSNSNCITWEGTNGEERKMTDPDEVARRWGERKSPNNYDKLSR 60

QY 367 ALRYTYDKNIMTVHPRESSMYKYPDDLPMSSYHGKRAKYKFDGIGAALQ----- 419
DB 61 ALRYTYDKNIMTVHPRESSMYKYPDDLPMSSYHGKRAKYKFDGIGAALQ----- 100

QY 420 -----PHAHPQKMFVAHPHPALPVTSSSFPAAPNPYNSPTG-GIYPN---TR 464
DB 101 MYRYQSDLYLPEYHPTKLNFVG--TPINPTNALSLESSHSSYSSWSPTCANITPSCGHVTH 158

QY 465 LPAHMPHSHLGYTY 478
DB 159 PHASHMSSHIGTY 172

RESULT 8
ETS6_DROME STANDARD: PRT: 475 AA.
AC P29776; Q9VP09;
DT 01-APR-1993 (rel. 25, Created)
DT 16-OCT-2001 (rel. 40, Last sequence update)
DT 15-JUN-2002 (rel. 41, Last annotation update)
DE DNA-binding protein D-ETS-6.
GN ETS21C OR ETS-6 OR CG2914.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Berkeley;
RX MEDLINE-20196006; PubMed-10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blake J.R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

```



```

RA Svltas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zhang X.H., Zhong F.N., Zhou W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of Drosophila melanogaster."
RA Science 287:2185-2195(2000).
RL
RL [3]
RL SEQUENCE OF 323-463 FROM N.A.
RL MEDLINE=91319397; PubMed=1713660;
RA Pribyl L.J., Watson D.R., Schulz R.A., Papas T.S.;
RA "D-elig, a member of the Drosophila ets gene family: sequence,
RA expression and evolutionary comparison."
RA Oncogene 6:1175-1183(1991).
RL
RL [4]
RL SEQUENCE OF 298-449 FROM N.A., FUNCTION, TISSUE SPECIFICITY, AND
RL DEVELOPMENTAL STAGE.
RA STRAIN=Canton-S; TISSUE=Larva;
RX MEDLINE=92249640; PubMed=1577186;
RA Chen T., Buntling M., Karim F.D., Thummel C.S.;
RA "Isolation and characterization of five Drosophila genes that encode
RA an ets-related DNA binding domain."
RL Dev. Biol. 151:176-191(1992).
CC
CC -1- FUNCTION: MAY HAVE A ROLE IN GERMLINE DEVELOPMENT.
CC
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC
CC -1- TISSUE SPECIFICITY: UNIFORM DISTRIBUTION THROUGHOUT EMBRYONIC
CC DEVELOPMENT, WITH SLIGHTLY HIGHER EXPRESSION IN POLE CELLS.
CC
CC -1- DEVELOPMENTAL STAGE: EXPRESSED THROUGHOUT DEVELOPMENT WITH LOWER
CC LEVELS DURING LARVAL DEVELOPMENT.
CC
CC -1- SIMILARITY: BELONGS TO THE ETS FAMILY.
CC
CC -1- SIMILARITY: CONTAINS 1 POINTED (PNT) DOMAIN.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X68259; CAA48327.1; -
DR EMBL: AF0033758; AAF56638.1; -
DR EMBL: X58481; CAA41390.1; ALT_INIT.
DR EMBL: M88471; AAC34199.1; -
DR PIR: S24300; S24300.
DR PIR: S28822; S28822.
DR PIR: S37616; S37616.
DR HSP: 000422; JAMC.
DR TRANSFAC: T02085; -
DR FLYBase: FBgn0004510; Ets97D.
DR InterPro: IPR000418; Ets.
DR InterPro: IPR002341; HSF_ETs.
DR InterPro: IPR003118; SAM_PNT.
DR Pfam: PF001178; Ets; 1.
DR Pfam: PF02198; SAM_PNT; 1.
DR PRINTS: PR00454; ETSDOMAIN.
DR SMART: SM00413; Ets; 1.
DR SMART: SM00251; SAM_PNT; 1.
DR PROSITE: PS00345; ETS_DOMAIN_1; 1.
DR PROSITE: PS00346; ETS_DOMAIN_2; 1.
DR PROSITE: PS00061; ETS_DOMAIN_3; 1.
DR PROSITE: PS00061; ETS_DOMAIN_3; 1.
KW DNA-binding; Nuclear Protein.
FT DOMAIN 186 269 POINTED.
FT DNA_BIND 346 425 ETS_DOMAIN.
FT CONFLICT 454 454 L -> V (IN REF. 2).
SQ SEQUENCE 464 AA; 52658 MW; 258A1F7C8D427A2 CRC64;
Query Match 17.6%; Score 455; DB 1; Length 464;
Best Local Similarity 32.1%; Pred. No. 1.2e-24;
Matches 113; Conservative 60; Mismatches 95; Indels 84; Gaps 10;
QY 73 NQVNGSINSPPDCSVAGKMKVSSDNVGNMGYSMEKRIHP-----PNTTNNRRY 125

```

```

Db 145 SOKSSSSSEPIKTPKRMKREDESEB-----EGKDVAVLNMWLDKFKKREQIRL 196
QY 126 IYPAPPTLMSTDHVMQEMAVKEYGLDPVLDLLPONDGKELCKMTXDDFORLPSTVA 185
Db 197 KIPPEANEMTHAHVYTWLEMAVKOFELVGINSDMO -MGDELCAHTEEFNOKLPRDG 255
QY 186 DILLSHLYLRE--TPLPHLTSDVDKALONSPRLMAR---MTGATFLFPMTSVYP 238
Db 266 NIFWTHLQLEKCNFVSVYHKAERQRPKQ--PLTMSANSTSTNSGSLSL----- 305
QY 239 EATQNTITRPDLPTQARRSAWTSLSHPQSKATQPPSSSTYVKTEDQRPDLPIYLPT 298
Db 306 --EORI-----MKRSYQSVKSSDSVSTSSMNPSTYT----- 337
QY 299 SSRLANPGSGOTLQWQFLELILSDSSNSCTWEGTNEEFKTDPDDEVARMRGERSKP 358
Db 338 ---IGSGNNGOVQWQFLELITLDCEDHVDLEWVTEGEFKLTPDRAALRGKRRKPA 394
QY 359 MNYDKLSRALRYDYDKNTKVPRESSMYKYPSPDLPMSSYHGKRYAVKFD 410
Db 395 MNYEKLSTALRYDYGDMSKY-----SGKRFAYKFD 426
RESULT 11
ETLA_XENLA STANDARD; PRT; 438 AA.
AC P18755;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE C-ETS-1A protein.
GN ETS-1A.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
OC Xenopodinae; Xenopus.
RX NCBI_TaxID=8355;
[1]
RP SEQUENCE FROM N.A.
RC TISSUE=Ovary;
RX MEDLINE=90384849; PubMed=2205841;
RA Stiegler P., Wolff C.M., Baltzinger M., Hirtzlin J., Senan F.,
RA Meyer D., Ghysdael J., Stehelin D., Belfort N., Remy P.;
RA "Characterization of Xenopus laevis cDNA clones of the c-ets-1 proto-
RA oncogene."
RL Nucleic Acids Res. 18:5298-5298(1990).
CC
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; C-ETS-1A (SHOWN HERE) AND C-
CC ETS-1A'; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC
CC -1- SIMILARITY: BELONGS TO THE ETS FAMILY.
CC
CC -1- SIMILARITY: CONTAINS 1 POINTED (PNT) DOMAIN.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X52692; CAA36919.1; -
DR PIR: S11225; S11225.
DR HSP: P14921; 2STT.
DR TRANSFAC: T02040; -
DR InterPro: IPR000418; Ets.
DR InterPro: IPR002341; HSF_ETs.
DR InterPro: IPR003118; SAM_PNT.
DR Pfam: PF001178; Ets; 1.
DR Pfam: PF02198; SAM_PNT; 1.
DR PRINTS: PR00454; ETSDOMAIN.
DR SMART: SM00413; Ets; 1.
DR SMART: SM00251; SAM_PNT; 1.

```

```

DR PROSITE: PS00345; ETS_DOMAIN_1; 1.
DR PROSITE: PS00346; ETS_DOMAIN_2; 1.
DR PROSITE: PS00347; ETS_DOMAIN_3; 1.
KW DNA-binding; Nuclear protein; Alternative splicing; Phosphorylation.
FT DOMAIN 51 134 POINTED.
FT DNA_BIND 332 412 ETS_DOMAIN.
FT MOD_RES 153 153 PHOSPHORYLATION (BY CAMK) (POTENTIAL).
FT MOD_RES 279 279 PHOSPHORYLATION (BY CAMK) (POTENTIAL).
FT VARSPLIC 1 203 MISSING (IN ISOFORM C-ETS-1A).
SQ SEQUENCE 438 AA; 50267 MW; B97A8EDF2DB51FA CRC64;

Query Match
Best local similarity 30.2%; Pred. No. 6,4e-24;
Matches 115; Conservative 50; Mismatches 102; Indels 114; Gaps 9;

QY 119 TTNERRVVPADPTLMSTDVHROMLEMAVKEVGLPVVDILLFQNDKELCKMTKDDFOR 178
DB 55 TKEQQRGLGIPIDREWTDHVRWVMAVNEFTLKGVDFOKF-CMGGAALCALGKECFLE 113
QY 179 LTFSTYADILSHLHLYREPLPHLTSDDYKALQN-----SRIMHARTGGATFI 230
DB 114 LAPDFVDILMEHLLELLOKSKQYOTSEITPAYPESRYTSDFYSYGIERRACVPESEFS 173
QY 231 FPN-----TSVPEAFORTTRDPLPYEQ----- 254
DB 174 EPEFTIESYOTLHPISSEILS---LKENYDPLGLRDPLOPESLOGDIYFTTKQEVYTP 230
QY 255 -----ARRSAWTSHPPTOSKATOPSSS-----TVPK-----TEDORP 287
DB 231 DNMCIGRISGRKLGGESFESIESHDSOBLTQSWSSQSSYNLQVNPVSDFSDSDYPR 250
QY 288 QLDPLYILG-----PSSKLANPGSQIOLWQFLLELSDSSNSNC 328
DB 291 AMSHSGKGFKNQYVDRAELNKKDPVIPAALAGYSGPILWQFLLELTDKSCQSP 350
QY 329 ITWEGTNGEFTMDPEVARRMGERKSKPMNMDKLSRLRYRYDKNIMTKVHPRESSMW 388
DB 351 ISWTGGEWFKLSDPEVARRMKRRKMKRMYEKLSRGLRYTDNIIHKT----- 402
QY 389 KYPSPDLPMSSYHGKRYAYKF 409
DB 403 -----AGKRYVRF 411

RESULT 12
GABA_HUMAN STANDARD; PRT; 454 AA.
AC Q06546; Q12939;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE GA binding protein alpha chain (GABP-alpha subunit) (Transcription factor E47F1-60) (Nuclear respiratory factor-2 subunit alpha).
GN E47F1A OR GABPA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93180783; PubMed=8441384;
RA Watanabe H., Sawada J.-I., Yano K.-I., Yamaguchi K., Goto M.,
RA Handa H.;
RT "cDNA cloning of transcription factor E47F1 subunits with Ets and
RT notch motifs";
RL MOL. Cell. Biol. 13:1385-1391(1993).
RP [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=95097980; PubMed=7799916;
RA Gudneja S., Virbasius J.V., Scarpulla R.C.;
RT "Four structurally distinct, non-DNA-binding subunits of human
RT nuclear respiratory factor 2 share a conserved transcriptional
RT activation domain.";
```

```

RL MOL. Cell. Biol. 15:102-111(1995).
CC -1- FUNCTION: TRANSCRIPTION FACTOR CAPABLE OF INTERACTING WITH PURINE
CC RICH REPEATS (GA REPEATS). NECESSARY FOR THE EXPRESSION OF THE
CC ADENOVIRUS E4 GENE.
CC -1- SUBUNIT: HETEROTETRAMER OF TWO ALPHA AND TWO BETA SUBUNITS.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE ETS FAMILY.
CC -1- SIMILARITY: CONTAINS 1 POINTED (PNT) DOMAIN.
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb.ch).
DR EMBL: D13318; BAA02575.1; -
DR EMBL: U13044; AAA05706.1; -
DR HSSP: Q00422; IAWC.
DR TRANSFAC: T01390; -
DR Genew: HGNC:4071; GABPA.
DR MIM: 600609; -
DR InterPro: IPR000418; Ets.
DR InterPro: IPR002341; HSF-ETS.
DR InterPro: IPR003118; SAM_PNT.
DR Pfam: PF00178; Ets; 1.
DR Pfam: PF02198; SAM_PNT; 1.
DR PRINTS: PR00454; ETSDOMAIN.
DR SMART: SM00413; ETS; 1.
DR SMART: SM00251; SAM_PNT; 1.
DR PROSITE: PS00345; ETS_DOMAIN_1; 1.
DR PROSITE: PS00346; ETS_DOMAIN_2; 1.
DR PROSITE: PS00347; ETS_DOMAIN_3; 1.
KW Transcription regulation; DNA-binding; Nuclear protein.
FT DOMAIN 170 251 POINTED.
FT DNA_BIND 320 400 ETS_DOMAIN.
FT CONFLICT 289 290 SS -> V (IN REF. 2).
FT CONFLICT 440 440 A -> V (IN REF. 2).
SQ SEQUENCE 454 AA; 51295 MW; 1AF2ABBC79191DD CRC64;

Query Match
Best local similarity 33.0%; Pred. No. 7.3e-24;
Matches 112; Conservative 48; Mismatches 94; Indels 85; Gaps 13;

QY 111 KHIPPNTTNER-----RVIVPADPTLMSTDVHROMLEMAVKEVGLPD 154
DB 150 KHTTTSDETSQVTRMAALEGYRKEGRGLGIPYPIOMSTIDQVLAHVVYWKKEFSMTD 209
QY 155 VDILFQNDKELCKMTKDDFORLTPSYNADILSHLHLYRETPPLHITS-----DVD 209
DB 210 IDLTTL-NISGRELSLQNDQFQRPV--RGETLMSLELRKYVALASQEQQNEIIVTD 266
QY 210 KALQNSPRLMHARNGAFTIPNYSVYEAFQRTTRDPLYEQARRSAMTSHSPTOS 269
DB 267 QPVQ-----IIPASVQSAT-----PTPI 284
QY 270 KATOPSSVTPKEDORPOLDYPOLIGPTSSRLANPGSGIOLWQFLLELSDSSNSNCI 329
DB 285 KYIN-SSAKAAVQ-RAPISSEDNSSP-GNRTGN--NQIOLWQFLLELTDKARDCT 339
QY 330 TWEGTNGEFTMDPEVARRMGERKSKPMNMDKLSRLRYRYDKNIMTKVHPRESSMWK 389
DB 340 SWVGDGEFKLQNPFLVAQKQGRNKKPMMNTEKLSRALRYYYDDDMICKVQ-GKRFYK 398
QY 390 YPSDLPYMSYHG-----KRYAYKDFHGIAQ 416
DB 399 FVCDLKTLLIGSABELNRLVTECEQKILA-KMQLHGIAQ 436

RESULT 13
GABA_MOUSE STANDARD; PRT; 454 AA.
ID GABA_MOUSE
```

AC 000422;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE GA binding protein alpha chain (GAPB-alpha subunit).
 GN GAPB1A OR GAPB.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91343912; PubMed=1876836;
 RA Lamarco K., Thompson C.C., Byers B.P., Walton E.M., McKnight S.L.;
 RT "Identification of Ets-and notch-related subunits in GA binding
 RT protein";
 RL Science 253:789-792(1991).
 RN [2]
 RP X-RAY CRYSTALLOGRAPHY (2.15 ANGSTROMS) OF 320-320.
 RX MEDLINE=98128030; PubMed=9461436;
 RA Batchelor A.H., Piper D.E., de la Brousse F.C., McKnight S.L.,
 RA Wolberger C.;
 RT "The structure of GAPBalpha/beta: an ETS domain-ankyrin repeat
 RT heterodimer bound to DNA";
 RL Science 279:1037-1041(1998).
 CC - FUNCTION: TRANSCRIPTION FACTOR CAPABLE OF INTERACTING WITH PURINE
 CC - RICH REPEATS (GA REPEATS).
 CC - SUBUNIT: HETEROTETRAMER OF TWO ALPHA AND TWO BETA SUBUNITS.
 CC - SUBCELLULAR LOCATION: Nuclear.
 CC - TISSUE SPECIFICITY: UBQUITOUS.
 CC - SIMILARITY: BELONGS TO THE ETS FAMILY.
 CC - SIMILARITY: CONTAINS 1 POINTED (PNT) DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: M74515; AAA53030.1; -
 DR PIR: A40858; A40858.
 DR PDB: 1AWC; 18-MAR-98.
 DR TRANSFAC: T00298; -
 DR TRANSFAC: T01402; -
 DR MGD: MGI:95610; Gabpa.
 DR InterPro: IPR000418; Ets.
 DR InterPro: IPR002341; HSF_ET5.
 DR InterPro: IPR003118; SAM_PNT.
 DR Pfam: PF00178; Ets; 1.
 DR Pfam: PF02198; SAM_PNT; 1.
 DR PRINTS: PR00454; ETSDOMAIN.
 DR SMART: SM00413; ETS; 1.
 DR SMART: SM00251; SAM_PNT; 1.
 DR PROSITE: PS00345; ETS_DOMAIN_1; 1.
 DR PROSITE: PS00346; ETS_DOMAIN_2; 1.
 DR PROSITE: PS50061; ETS_DOMAIN_3; 1.
 KM Transcription regulation; DNA-binding; Nuclear protein; 3D-structure.
 FT DNA_BIND 320 400 ETS-DOMAIN.
 FT SEQUENCE 454 AA; 51363 MW; 76BC47EB2F563AAND CRC64;
 SQ
 Query Match 17.1%; Score 441.5; DB 1; Length 454;
 Best Local Similarity 28.6%; Pred. No. 1e-23;
 Matches 126; Conservative 66; Mismatches 127; Indels 119; Gaps 17;
 QY 15 DQSLFCAVSPHLAKTEMASSSEYQGSKSPRPVQDWTLSQPARVITKKECNPNQ 74
 DB 78 DRSLFDG-----VKTDGYQLSVQVISTYQGMKPKL---NILEIVTAETAEVEVIDPD- 127
 QY 75 VNGSRNPPDQCSVAKGKMMSSSDNVMNMGNSYMEKHIPPNNMTYMER----- 123

DB 128 ---AHHAFAEAHVEAQTLDGT-----KHITTSDETSQVTRMAALEGY 173
 QY 124 -----RVIVPADPTLWSTDHVROWLEMAVKEYGLPDVDILFQNDIGKELCMTKDDFOR 178
 DB 174 RKEGEGIPYDPIRMSTDOYLHWVVMWYMKFEFMTIDITLTL-NISGEELCSLNGEDPFQ 232
 QY 179 LTPSYMDILSHLYLEPPLPHLTS-----DVKALQNSPRLMARNTGATFTFPN 233
 DB 233 RVP--RGEILSHLELLRKVYLAQEQOMNEIVIDQVQILP----- 273
 QY 234 TSVYPEARQRTTRPDLPEQARRSAMTSHSPQSKATQSSSVKPTEDQRPQLDYQ 293
 DB 274 -ASVPAPT-----PTTKYINSAAKA-----KVQRP-R 301
 QY 294 ILGPTSSRLANPG-----SCQIQLOMFLLELLSDSSNSNCITWGTNGEFTKMTDPDEVAR 348
 DB 302 ISGEDRS---SPGNRTGNGGQIQLOMFLLELLTBKARDCLSWGDEBEFTLNDELVAQ 358
 QY 349 RWGERKSKPNMYDKLSRALRYTDKNIMTVHPPESSMYRPPSDLPYMSYHG----- 402
 DB 359 KWGQRKNKPTMYNRYEKLRSALRYTDGDMICKVQ--GKRFVYFVCDLKTLLIGYSAAELNRL 417
 QY 403 -----KRYAKFDFHGIAQ 416
 DB 418 VICEQKKLA-RMQLHGIAQ 436
 RESULT 14
 ETS2_CHICK STANDARD; PRT; 479 AA.
 AC P10157;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 16-MAR-1989 (Rel. 10, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE C-ETS-2 protein.
 GN ETS2.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archaeopteryx; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Fibroblast;
 RX MEDLINE=88283637; PubMed=3293999;
 RA Bouloukos K.E., Pogonue P., Begue A., Galibert F., Gesquiere J.C.,
 RA Stehelin D., Ghysdael J.;
 RT "Identification in chickens of an evolutionarily conserved cellular
 RT ets-2 gene (c-ets-2) encoding nuclear proteins related to the
 RT products of the c-ets proto-oncogene";
 RL EMBO J. 7:697-705(1988).
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- SIMILARITY: BELONGS TO THE ETS FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 POINTED (PNT) DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: X07202; CAA30178.1; -
 DR PIR: S00386; TYCHE2.
 DR HSSP: P14921; 2STT.
 DR TRANSFAC: T00116; -
 DR InterPro: IPR000418; Ets.
 DR InterPro: IPR002341; HSF_ET5.
 DR InterPro: IPR003118; SAM_PNT.
 DR Pfam: PF00178; Ets; 1.
 DR Pfam: PF02198; SAM_PNT; 1.
 DR PRINTS: PR00454; ETSDOMAIN.

DR SMART: SM00413; ETS: 1.
 DR SMART: SM00251; SAM_PNT: 1.
 DR PROSITE; PS00345; ETS_DOMAIN_1; 1.
 DR PROSITE; PS00346; ETS_DOMAIN_2; 1.
 DR PROSITE; PS00346; ETS_DOMAIN_3; 1.
 KW Proto-oncogene: DNA-binding; Nuclear protein.
 FT DOMAIN 90 173 POINTED.
 FT DNA_BIND 373 453 ETS-DOMAIN.
 SQ SEQUENCE 479 AA; 54540 MW; 91BCD5206972E867 CRC64;
 Query Match 17.0%; Score 441; DB 1; Length 479;
 Best Local Similarity 31.1%; Pred. No. 1,2e-23;
 Matches 118; Conservative 44; Mismatches 108; Indels 110; Gaps 12;
 QY 119 TTNERRVIVPADPTLMSTHVRQMLEMAVKEYGLPDVDILLFQNDGKELCKMTKDDFOR 178
 DB 94 TREQQRIGIPNNWLTEDVHCQMLAMATNEFSLAVNVAHQFL-MSGODLCNIGKERFLE 152
 QY 179 LTPSYNADILLSHLYLRE-----TPPLHLSDS-----DVD----- 209
 DB 153 LAPDYGDILMEHLQMIKDSQKTDQVYESHLTSVPHWNNNSLTIVAVDQTPYGIOM 212
 QY 210 ---KALQ-NSPLMH---ARNTGATF-----IFPNTSVYPEATORTTTPDL--- 250
 DB 213 PGYKALSYPKPVLSDICQSTGTPMLSPEDQFSLFPTQVDAVSNTCTVADFTRSN 272
 QY 251 -----PYE--QARRKAWTSHS-----PTQSK----- 270
 DB 273 LNLILDSGKLRHESSESGAESSESSDMLQSNSSSLVDLQRYPSYSEFEDDCSQSL 332
 QY 271 -ATOPSSSTVPKTEDORPOLDPYQIIGPTSSRLANPGSGQIOLMOFLLELSSSSNCT 329
 DB 333 CASKPTMSFEDYIQDSDEYQSKPYIPAILAGFGSGTQIMQFLLELTKSCQSFI 352
 QY 330 TWEGTNGEFKMDPDVEARRWGRKSKPNNYDKLSRALRYVYDKNIMTKVHPRESSMYK 389
 DB 393 SWTGDMGEFLADPDDEVARWGRKKNPKMYNKEKLSRGLATYYDKNLIHKT----- 443
 QY 390 YPSDLPYMSYHGKRYAYKF 409
 DB 444 -----SGKRYRFR 452
 RESULT 15
 ETS1_HUMAN STANDARD: PRT; 441 AA.
 ID ETS1_HUMAN
 AC P14921;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE C-ets-1 protein (p54).
 GN ETS1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=89042086; PubMed=2847145;
 RA Watson D.K., McWilliams M.J., Lapis P., Lautenberger J.A.,
 RA Scheinfeist C.W., Papas T.S.,
 RA "Mammalian ets-1 and ets-2 genes encode highly conserved proteins,"
 RL Proc. Natl. Acad. Sci. U.S.A. 85:7862-7866(1988).
 RN [3]
 RP STRUCTURE BY NMR OF 320-415.
 RX MEDLINE=96097120; PubMed=8521493;

RA Werner M.H., Clore G.M., Fisher C.L., Fisher R.J., Trinh L.,
 RA Shiloach J., Gronenborn A.M.;
 RT "The solution structure of the human ETS1-DNA complex reveals a novel
 RT mode of binding and true side chain intercalation,"
 RL Cell 83:761-771(1995).
 CC 1- SUBCELLULAR LOCATION: Nuclear.
 CC 1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; C-ETS-1A (SHOWN HERE) AND C-ETS-
 CC 1B; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC 1- PTM: ISOFORMS OF ETS-1 ARE ALSO GENERATED BY DIFFERENTIAL
 CC PHOSPHORYLATION.
 CC 1- DISEASE: ETS IS RESPONSIBLE FOR ERYTHROBLAST AND FIBROBLAST
 CC TRANSFORMATION. THE JUXTAPPOSITION OF THE INTERFERON C-ETS-1
 CC PROTO-ONCOGENE MAY BE INVOLVED IN THE PATHOGENESIS OF HUMAN
 CC MONOCYTIC LEUKEMIA.
 CC 1- SIMILARITY: BELONGS TO THE ETS FAMILY.
 CC 1- SIMILARITY: CONTAINS 1 POINTED (PNT) DOMAIN.
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 DR EMBL; X14798; CAA32904.1; -
 DR EMBL; X14798; CAA32903.1; -
 DR EMBL; J04101; AAA52410.1; -
 DR PIR; A32066; TVHUET.
 DR PIR; S10086; S10086.
 DR PDB; 2STW; 12-MAR-97.
 DR PDB; 2STW; 12-MAR-97.
 DR TRANSFAC; T00112; -
 DR TRANSFAC; T01400; -
 DR GeneW; HGNC:3488; ETS1.
 DR MIM; 164720; -
 DR InterPro; IPR000418; Ets.
 DR InterPro; IPR002341; HSF-ETS.
 DR InterPro; IPR003118; SAM_PNT.
 DR Pfam; PF00178; Ets; 1.
 DR Pfam; PF02198; SAM_PNT; 1.
 DR PRINTS; PR00454; ETSDOMAIN.
 DR SMART; SM00413; ETS; 1.
 DR SMART; SM00251; SAM_PNT; 1.
 DR PROSITE; PS00345; ETS_DOMAIN_1; 1.
 DR PROSITE; PS00346; ETS_DOMAIN_2; 1.
 DR PROSITE; PS00346; ETS_DOMAIN_3; 1.
 KW Proto-oncogene: DNA-binding; Nuclear protein; Alternative splicing;
 KW Phosphorylation; 3D-structure.
 FT DOMAIN 53 136 POINTED.
 FT DNA_BIND 335 415 ETS-DOMAIN.
 FT MOD_RES 38 38 PHOSPHORYLATION (BY MAPK) (BY
 FT SIMILARITY).
 FT VARSPIC 244 330 MISSING (IN ISOFORM C-ETS-1B).
 SQ SEQUENCE 441 AA; 50408 MW; 3B66BCC464B393FB CRC64;
 Query Match 17.0%; Score 440.5; DB 1; Length 441;
 Best Local Similarity 30.1%; Pred. No. 1,1e-23;
 Matches 114; Conservative 54; Mismatches 102; Indels 109; Gaps 8;
 QY 119 TTNERRVIVPADPTLMSTHVRQMLEMAVKEYGLPDVDILLFQNDGKELCKMTKDDFOR 178
 DB 57 TREQQRIGIPKDPROWTETTHVRDWWMAVNEFSLKGVDFQKE-CMNGALCALGRDCEFL 115
 QY 179 LTPSYNADILLSHLYLRETPPLHLSDDVDKALQNSP-----RLMHARNTGATF 229
 DB 116 LAPDYGDILMEHLQMIKDSQKTDQVYESHLTSVPHWNNNSLTIVAVDQTPYGIOM 175
 QY 230 IFPN-----TSVPEATORTI----- 244
 DB 176 SEPSFTTESYQTLHPISSELSLKYENDYPSVILRDLQDTPLQNDYPAIKQEVVTPDN 235
 QY 245 -----TTRPDLPEQARRS--AMTSHSPTQSKATQPPSSSTVPK-----TED----- 284

```
Db 236 MCMGRTSRGKLGQDSFESTIESYDSCDRLTQSWSSQSSFNSLQRPSPYDSFDSEDYPAL. 295
QY 285 -----QRPQLDPYQILGPTSSRLANPGSGQIQIMQFLELLSDSSNSNCIT 330
Db 296 PNHKPKGTFFKDYVRDRADLNKDKVITPAALAGYTGSGPIQLMQFLELLTDKSCQSFIS 355
QY 331 WEGTNGEFFKATDPDEVARRWGERKSKPNMNYDKLSRALRYDYDKNIMTKVHPRESSMYKY 390
Db 356 WTGDGMEFEKLSDPDEVARRWGKRKNKPKMNYEKLRSGLRYDYDKNIIHKT----- 405
QY 391 PSDLPYMSYHGKRYAKKF 409
Db 406 -----AGKRYRYRF 414
```

Search completed: November 9, 2002, 16:51:18
Job time : 22.0958 secs

THIS PAGE BLANK (USPTO)

Db	69	ARVITIKMCNCPSQVGNSSNPDECSYKNGGAGVGS	PPVIGKSYSSYMEEKHVPPNMTTN	128	
Qy	122	ERRVYVPADPPLMSTDHRYQWLENAVYKEYGL	PVDVILLPQNIIDKELCCKATKDDFORLTP	181	
Db	129	ERRVYVPADPPLMSTDHRYQWLENAVYKEYGL	PVDVILLPQNIIDKELCCKATKDDFORLTP	188	
Qy	182	SYNADILLSHLYLEPFLPHLTSDDDVKKALQNS	PRMLHARNTGGATFIPNPTSVEAT	241	
Db	189	SYNADILLSHLYLEPFLPHLTSDDDVKKALQNS	PRMLHARNTGGATFIPNPTSVEAT	248	
Qy	242	QRITTRPDLPEQARRSAMTSHSPT-QSKATOPS	STVKTEDQRPOLDPYQILGPTSS	300	
Db	249	QRITTRPDLPEQARRSAMTSHSPT-QSKATOPS	STVKTEDQRPOLDPYQILGPTSS	308	
Qy	301	RLANGSQOIQDLMQFLELLSDSSNSNCITWEG	NGEKKMTDPEVARRMGERKSRNMN	360	
Db	309	RLANGSQOIQDLMQFLELLSDSSNSNCITWEG	NGEKKMTDPEVARRMGERKSRNMN	368	
Qy	361	YDKLSRALRYYYDKNIMTKVHPPESSMYKYP	SDLPYMSYHGKRAYAKFDFHIGQALQ	420	
Db	369	YDKLSRALRYYYDKNIMTKVHPPESSMYKYP	SDLPYMSYHGKRAYAKFDFHIGQALQ	408	
Qy	421	-----	HAHQKKNFYAPHPHPALPYTSSSFPAAPN	YNNMSPGIGIT	460
Db	409	HPPESSLYKTYSDLPYMSYHAHPQKKNFY	SPHPALPYTSSSFPAAPN	YNNMSPGIGIT	468
Qy	461	PNTRLPAAHMPSHLGTYT	478		
Db	469	PNTRLPAAHMPSHLGTYT	486		
RESULT 2					
ID	Q80UD0	PRELIMINARY;	PRT;	451	AA.
AC	Q80UD0				
DT	01-MAR-2002	(TREMBlrel. 20, Created)			
DT	01-MAR-2002	(TREMBlrel. 20, Last sequence update)			
DT	01-JUN-2002	(TREMBlrel. 21, Last annotation update)			
DE	Erg isoform C-1-1.				
OS	Gallus gallus (Chicken).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;				
OC	Gallus.				
OX	NCBI_TaxID=9031;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=20351415; PubMed=10893254;				
RA	Iwamoto M., Higuchi Y., Koyama E.,				
RA	Yeh H., Rosenbloom J., Pacifici M.;				
RL	"The role of ERK (ets related gene) in cartilage development.";				
RL	OSTEOARTHRITIS Cartilage 9:S41-S47(2001).				
DR	EMBL; AY056561; AAL40889.1; "				
DR	InterPro: IPR000418; Ets.				
DR	InterPro: IPR002341; HSF-ETS.				
DR	InterPro: IPR003118; SAM_PNT.				
DR	Pfam: PF001178; Ets. 1.				
DR	Pfam: PF021198; SAM_PNT. 1.				
DR	PRINTS; PR00454; ETSDOMAIN.				
DR	SMART; SM00413; ETS. 1.				
DR	SMART; SM00251; SAM_PNT. 1.				
DR	PROSITE; PS00345; ETS_DOMAIN_1; UNKNOWN_1.				
DR	PROSITE; PS00346; ETS_DOMAIN_2; UNKNOWN_1.				
DR	PROSITE; PS50061; ETS_DOMAIN_3; 1.				

DR	PROSITE: PS50105; SAM_DOMAIN: 1.	51002 MM; 3ECC4B8CA615AA31 CRC64:
SEQ	SEQUENCE	451 AA; 51002 MM; 3ECC4B8CA615AA31 CRC64:
Query Match	87.4%; Score 2262.5; DB 13; Length 451;	
Best Local Similarity	86.3%; Pred. No. 1.7e-166;	
Matches 430; Conservative	0; Mismatches 1; Indels 67; Gaps 3	
QY	1 MASTIKELSVSEDOSEFECAYGSHLAKTEMTASSSEYQOTSKMSRPVQOQMLSP	60
DB	1 MASTIKELSVSEDOSEFECAYGSHLAKTEMTASSSEYQOTSKMSRPVQOQMLSP	60
QY	61 PARVIKMECNQVNGSNSPDCSVAKGKMWSSDNVGNNGYSYMEKHIIPPNMT	120
DB	61 PARVIKMECNQVNGSNSPDCSVAKGKMWSSDNVGNNGYSYMEKHIIPPNMT	120
QY	121 NERRVIVPADPTLMSTDHVRQMLENAVKEGLPDVDILFQNDIKELCKMRDDFORLT	180
DB	121 NERRVIVPADPTLMSTDHVRQMLENAVKEGLPDVDILFQNDIKELCKMRDDFORLT	180
QY	181 PSYNDILLSHHTYREPRPLPHLTSDVDYDKALQNSBRLMAHNTGATFIPTSYVPEA	240
DB	181 PSYNDILLSHHTYREPRPLPHLTSDVDYDKALQNSBRLMAHNTGATFIPTSYVPEA	240
QY	241 TORITRPDLPEQARRSMTSHSPQSKATOPSSSYPTKEDORPOLDPYQILGPTSS	300
DB	241 TORITRPDLPEQARRSMTSHSPQSKATOPSSSYPTKEDORPOLDPYQILGPTSS	300
QY	301 RLANPGSGOIQLMQRLLELSDSSNSNCITWEGTGEEFKMTDPDVAARWGERKSKPMNN	360
DB	274 RLANPGSGOIQLMQRLLELSDSSNSNCITWEGTGEEFKMTDPDVAARWGERKSKPMNN	333
QY	361 YDKLSRALRYYYDKKIMTKV-----HPSSSMYKYPSSDLPYMSY	400
DB	334 YDKLSRALRYYYDKKIMTKV-----HPSSSMYKYPSSDLPYMSY	393
QY	401 HGKRAVYKFDHGIAQALQPHAHPOKMFVAPHPALPVTSSFFRAANPYKNSSTGGIT	460
DB	394 -----HAHPQKMFVAPHPALPVTSSFFRAANPYKNSSTGGIT	433
QY	461 PNTRLPAAHMPSHLGTYT 478	
DB	434 PNTRLPAAHMPSHLGTYT 451	
RESULT 3		
Q920K8	PRELIMINARY; PRT; 463 AA.	
ID	Q920K8	
AC	Q920K8	
DT	01-DEC-2001 (TREMBLrel. 19, Created)	
DT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)	
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)	
DE	Erg protein.	
GN	Erg.	
OS	Mus musculus (Mouse).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
CC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.	
OX	NCBI_TaxID=10090;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RA	Ozawa R., Noguchi H., Taylor T.D., Takeda T., Hattori M., Sakaki Y.;	
RT	"Mus musculus Erg mRNA."	
RL	Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.	
DR	EMBL; AB073079; BAB69949.1; -	
DR	MED; MGI:95415; Erg.	
DR	InterPro; IPR000418; ETS.	
DR	InterPro; IPR002341; HSF_ET5.	
DR	InterPro; IPR003118; SAM_PNT.	
DR	Pfam; PF00178; Ets; 1.	
DR	Pfam; PF02198; SAM_PNT; 1.	
DR	PROSITE; PS00345; ETS_DOMAIN_1; UNKNOWN_1.	
DR	PROSITE; PS00346; ETS_DOMAIN_2; UNKNOWN_1.	
DR	PROSITE; PS50061; ETS_DOMAIN_3; 1.	
Q0	SEQUENCE 463 AA; 51971 MW; 510D2BB76663D4A4 CRC64;	

Query Match 84.5%; Score 2186; DB 11; Length 463;
 Best Local Similarity 83.1%; Pred. No. 1.6e-162;
 Matches 413; Conservative 11; Mismatches 11; Indels 62; Gaps 3;

QY 2 ASTKEALSVSEDOSEFECAYGSPHLAKTEMTASSSESYQTSKMSPRVPOQDMLSQP 61
 DQ 9 AAHTKEALSVSEDOSEFECAYGSPHLAKTEMTASSSESYQTSKMSPRVPOQDMLSQP 68
 QY 62 ARVITKMECNNOVNGSNPDDCSVAKGKMWSSSDNVGNGNSYMEKHPPPNMTN 121
 DQ 69 ARVITKMECNNOVNGSNPDDCSVAKGKMWSSSDNVGNGNSYMEKHPPPNMTN 128
 QY 122 ERRVIVADPTLMSTIDHVRQWLEMAVKEYGLPDVDILLFQNDIGKELCKMTKDDFORLTP 181
 DQ 129 ERRVIVADPTLMSTIDHVRQWLEMAVKEYGLPDVDILLFQNDIGKELCKMTKDDFORLTP 188
 QY 182 SYNDILSHLHYLRETPPLHLTSDVDKALQNSPRLMHARNTGGATIFPNTSYPEAT 241
 DQ 189 SYNDILSHLHYLRETPPLHLTSDVDKALQNSPRLMHARNTGGATIFPNTSYPEAT 248
 QY 242 QRTTRPDPYEQARSAWTSHPQSKATOPSSSTVPTKEDORPOLDPYQILGPTSSR 301
 DQ 249 QRTTRPDPYEQARSAWTSHPQSKATOPSSSTVPTKEDORPOLDPYQILGPTSSR 286
 QY 302 LANSGQIQLOMOLFLELLSDSSNSNCITWGTNGEFTKMDPDEVARRMGRKSKPMNMY 361
 DQ 287 LANSGQIQLOMOLFLELLSDSSNSNCITWGTNGEFTKMDPDEVARRMGRKSKPMNMY 346
 QY 362 DKLSBALRYYYDKNIMTKVHPPESSMYKYPDSLPYMSYHGKRYAYKDFHGIAQALOP 420
 DQ 347 DKLSBALRYYYDKNIMTKVHPPESSMYKYPDSLPYMSYHGKRYAYKDFHGIAQALOP 386
 QY 421 -----HAHPKMFVAPHPALPVTSSEFFAAPNPYNSPTGGIYLP 461
 DQ 387 PPESLHYPSDDLPMYSGYHAHPQKMFVSPHPALPVTSSEFFAAPNPYNSPTGGIYLP 446
 QY 462 NTRLPAAHMPSHLGTYY 478
 DQ 447 NTRLPAAHMPSHLGTYY 463

RESULT 4
 Q9W700 PRELIMINARY; PRT; 485 AA.
 ID Q9W700;
 AC Q9W700;
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
 DE Erg protein.
 GN ERG.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Amphibia; Batrachia; Anura; Mesodactylia; Pipridae; Pipridae;
 OC Xenopodinae; Xenopus.
 CX NCBL_TaxID=8355;
 RX NCBL_TaxID=8355;
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20099678; PubMed=10633861;
 RA Baltzinger M., Mager-Heckel A.M., Remy P.;
 RT "Xlerg: Expression and pattern and overexpression during development
 RT pleid for a role in endothelial cell differentiation.";
 RL Dev. Dyn. 216:420-433(1999).
 DR EMBL: AJ224125; CAB46566.1; -
 DR HSSP: A01543; IFLI
 DR InterPro: IPR000418; Ets.
 DR InterPro: IPR002341; HSF_ETS.
 DR InterPro: IPR003118; SAM_PNT.
 DR Pfam: PF00178; Ets; 1.
 DR Pfam: PF02198; SAM_PNT; 1.
 DR PRINTS: PR00454; ETSDOMAIN.
 DR SMART: SM00413; ETS; 1.
 DR SMART: SM00251; SAM_PNT; 1.

DR PROSITE; PS00345; ETS_DOMAIN_1; 1.
 DR PROSITE; PS00346; ETS_DOMAIN_2; 1.
 DR PROSITE; PS0061; ETS_DOMAIN_3; 1.
 SO SEQUENCE 485 AA; 54804 MW; F87BF6DAFBA279F CRC64.

Query Match 84.2%; Score 2180; DB 13; Length 485;
 Best Local Similarity 83.1%; Pred. No. 5.1e-162;
 Matches 409; Conservative 19; Mismatches 22; Indels 42; Gaps 4;

QY 7 EALSVSEDOSEFECAYGSPHLAKTEMTASSSESYQTSKMSPRVPOQDMLSQPARYTI 66
 DQ 16 EALSVSEDOSEFECAYGSPHLAKTEMTASSSESYQTSKMSPRVPOQDMLSQPARYTI 75
 QY 67 KMECPNOVNGSNPDDCSVAKGKMWSSSDNVGNGNSYMEKHPPPNMTNERRVY 126
 DQ 76 KMECPNOVNGSNPDDCSVAKGKMWSSSDNVGNGNSYMEKHPPPNMTNERRVY 135
 QY 127 VPADPTLMSTIDHVRQWLEMAVKEYGLPDVDILLFQNDIGKELCKMTKDDFORLTPSYND 186
 DQ 136 VPADPTLMSTIDHVRQWLEMAVKEYGLPDVDILLFQNDIGKELCKMTKDDFORLTPSYND 195
 QY 187 ILSHLHYLRETPPLHLTSDVDKALQNSPRLMHARNTGGATIFPNTSYPEATQRTYT 246
 DQ 196 ILSHLHYLRETPPLHLTSDVDKALQNSPRLMHARNTGGATIFPNTSYPEATQRTYT 255
 QY 247 RPDLPYEQARSAWTSHPQSKATOPSSSTVPTKEDORPOLDPYQILGPTSSRLANPG 306
 DQ 256 RPDLPYEQARSAWTSHPQSKATOPSSSTVPTKEDORPOLDPYQILGPTSSRLANPG 313
 QY 307 SGOIQLOMOLFLELLSDSSNSNCITWGTNGEFTKMDPDEVARRMGRKSKPMNMYDKLSR 366
 DQ 314 SGOIQLOMOLFLELLSDSSNSNCITWGTNGEFTKMDPDEVARRMGRKSKPMNMYDKLSR 373
 QY 367 ALRYYYDKNIMTKVHPPESSMYKYPDSLPYMSYHGKRYAYKDFHGIAQALOP----- 420
 DQ 374 ALRYYYDKNIMTKVHPPESSMYKYPDSLPYMSYHGKRYAYKDFHGIAQALOP----- 413
 QY 421 -----HAHPKMFVAPHPALPVTSSEFFAAPNPYNSPTGGIYNTNRLP 466
 DQ 414 MYKYPSELPMYSGYHAHPQKMFVAPHPALPVTSSEFFAAPNPYNSPTGGIYNTNRLP 473
 QY 467 AAHMPSHLGTYY 478
 DQ 474 AAHMPSHLGTYY 485

RESULT 5
 Q91XV5 PRELIMINARY; PRT; 455 AA.
 ID Q91XV5;
 AC Q91XV5;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, last annotation update)
 DE Vascular endothelial cell specific protein 14.
 GN VESPL4.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 CX NCBL_TaxID=10116;
 RX NCBL_TaxID=10116;
 RP SEQUENCE FROM N.A.
 RA Aoki T., Toyoda H., Nishimoto S., Tawara J., Ukai Y., Komurasaki T.;
 RT "Identification of VESPL4, a vascular endothelial cell specific
 RT protein.";
 RL Submitted (Aug-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB031088; BAB62744.1; -
 DR InterPro: IPR000418; Ets.
 DR InterPro: IPR002341; HSF_ETS.
 DR InterPro: IPR003118; SAM_PNT.
 DR Pfam: PF00178; Ets; 1.
 DR Pfam: PF02198; SAM_PNT; 1.
 DR PROSITE; PS00345; ETS_DOMAIN_1; UNKNOWN_1.

DR PROSITE: PS00346; ETS_DOMAIN_2; UNKNOWN_1.
 DR PROSITE: PS00061; ETS_DOMAIN_3; 1
 SQ SEQUENCE 455 AA; 51382 MM; B7E2A1564F88560D CRC64;
 Query Match 84.1%; Score 2177.5; DB 11; Length 455;
 Best Local Similarity 82.6%; Pred. No. 7.3e-162;
 Matches 412; Conservative 9; Mismatches 13; Indels 65; Gaps 4;

QY 1 MASTIEALSVSEDSQSLFECAYGSPHLAKTEMTASSSEYGGTSSKSPRPQODWLSQP 60
 |||
 DB 1 MASTIEALSVSKDSQSLFECAYGTPHLAKTEMTASSSSDYGGTSSKSPRPQODWLSQP 60
 |||
 QY 61 PAVTIKMECPNPNVNGSRNSPDDCSYAKGKAVSSSDNVMGYSGMEKHIPPMMTT 120
 |||
 DB 61 PAVTIKMECPNPNVNGSRNSPDDCSYAKGKAVSSSDNVMGYSGMEKHIPPMMTT 120
 |||
 QY 121 NERRVIVPADPTLMSTDHVRQWLEMAVKEYGLPDVILLFQNDIGKELCKMTKDDFQRLT 180
 |||
 DB 121 NERRVIVPADPTLMSTDHVRQWLEMAVKEYGLPDVILLFQNDIGKELCKMTKDDFQRLT 180
 |||
 QY 181 PSYNADILLSHLHYLRETPPLHLTSDVDKALQNSPRLMHARNTGATFTFPNTSVYPEA 240
 |||
 DB 181 PSYNADILLSHLHYLRETPPLHLTSDVDKALQNSPRLMHARNTGATFTFPNTSVYPEA 240
 |||
 QY 241 TQRTTRPDLPYEQARRSAMTSHSHT-OSKATQPSSTVKTEDQRPOLDPQIIGPTSS 299
 |||
 DB 225 -----DLPEPPRRSMTGSHHTPOSKAAPSPSIVPKTEDQRPOLDPQIIGPTSS 276
 |||
 QY 300 SRLANSGOIOLMOFLLELLSDSSNSNCITWEGTNGEFTKMTDDEVARRGERSKSPNM 359
 |||
 DB 277 SRLANSGOIOLMOFLLELLSDSSNSNCITWEGTNGEFTKMTDDEVARRGERSKSPNM 359
 |||
 QY 360 NYDKLSRALRYDYDKNIMTKVHPRESSMYKPSDLPYMSYHGKRRAYKDFHGIQALQ 419
 |||
 DB 337 NYDKLSRALRYDYDKNIMTKV-----HGKRRAYKDFHGIQALQ 376
 |||
 QY 420 P-----HAHPQKMFVAPHPALPYTSSSFPAAPNPYNSPTGCI 459
 |||
 DB 377 PHPESSLYKKPSDLPYMGSYHHPQKMFVAPHPALPYTSSSFPAAPNPYNSPTGCI 436
 |||
 QY 460 YPNTRLPAAMHPSHLGTYT 478
 |||
 DB 437 YPNTRLPAAMHPSHLGTYT 455

RESULT 6
 Q920K9 PRELIMINARY; PRT; 462 AA.
 AC Q920K9;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Erg protein.
 GN ERG.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_Taxid=10090;
 RN (1)
 RP SEQUENCE FROM N.A.
 RA Ozaava R., Noguchi H., Taylor T.D., Takeda T., Hattori M., Sakaki Y.;
 RT "Mus musculus Erg mRNA.";
 RL Submitted (OCU-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB073078; BAB69948.1; -;
 DR MGD; MGI:95415; Erg.
 DR InterPro: IPR000418; Ets.
 DR InterPro: IPR002341; HSF_ETS.
 DR InterPro: IPR003118; SAM_PNT.
 DR Pfam; PF00178; Ets; 1.
 DR InterPro: IPR002341; HSF_ETS.
 DR InterPro: IPR003118; SAM_PNT.
 DR Pfam; PF00178; Ets; 1.
 DR PROSITE; PS00345; ETS_DOMAIN_1; UNKNOWN_1.
 DR PROSITE; PS00346; ETS_DOMAIN_2; UNKNOWN_1.
 DR PROSITE; PS00061; ETS_DOMAIN_3; 1.

SQ SEQUENCE 462 AA; 52037 MM; 0EA419B53ECAA591 CRC64;
 Query Match 83.1%; Score 2149.5; DB 11; Length 462;
 Best Local Similarity 81.9%; Pred. No. 1.1e-159;
 Matches 408; Conservative 11; Mismatches 14; Indels 65; Gaps 4;

QY 2 ASITKEALSVSEDSQSLFECAYGSPHLAKTEMTASSSEYGGTSSKSPRPQODWLSQP 61
 |||
 DB 9 AAHIKEALSVSEDSQSLFECAYGTPHLAKTEMTASSSSDYGGTSSKSPRPQODWLSQP 68
 |||
 QY 62 ARVTIKMECPNPNVNGSRNSPDDCSYAKGKAVSSSDNVMGYSGMEKHIPPMMTTN 121
 |||
 DB 69 ARVTIKMECPNPNVNGSRNSPDDCSYAKGKAVSSSDNVMGYSGMEKHIPPMMTTN 128
 |||
 QY 122 ERRVIVPADPTLMSTDHVRQWLEMAVKEYGLPDVILLFQNDIGKELCKMTKDDFQRLT 181
 |||
 DB 129 ERRVIVPADPTLMSTDHVRQWLEMAVKEYGLPDVILLFQNDIGKELCKMTKDDFQRLT 188
 |||
 QY 182 SYNADILLSHLHYLRETPPLHLTSDVDKALQNSPRLMHARNTGATFTFPNTSVYPEAT 241
 |||
 DB 189 SYNADILLSHLHYLRETPPLHLTSDVDKALQNSPRLMHARNTGATFTFPNTSVYPEAT 231
 |||
 QY 242 QRTTRPDLPYEQARRSAMTSHSHT-OSKATQPSSTVKTEDQRPOLDPQIIGPTSS 300
 |||
 DB 232 -----DLPEPPRRSMTGSHHTPOSKAAPSPSIVPKTEDQRPOLDPQIIGPTSS 284
 |||
 QY 301 RLANPSGOIOLMOFLLELLSDSSNSNCITWEGTNGEFTKMTDDEVARRGERSKSPNM 360
 |||
 DB 285 RLANPSGOIOLMOFLLELLSDSSNSNCITWEGTNGEFTKMTDDEVARRGERSKSPNM 344
 |||
 QY 361 YDKLSRALRYDYDKNIMTKVHPRESSMYKPSDLPYMSYHGKRRAYKDFHGIQALQ 420
 |||
 DB 345 YDKLSRALRYDYDKNIMTKV-----HGKRRAYKDFHGIQALQ 384
 |||
 QY 421 P-----HAHPQKMFVAPHPALPYTSSSFPAAPNPYNSPTGCI 460
 |||
 DB 385 HPPESSLYKKPSDLPYMGSYHHPQKMFVAPHPALPYTSSSFPAAPNPYNSPTGCI 444
 |||
 QY 461 PNTRLPAAMHPSHLGTYT 478
 |||
 DB 445 PNTRLPAAMHPSHLGTYT 462

RESULT 7
 Q9W6Z9 PRELIMINARY; PRT; 456 AA.
 AC Q9W6Z9;
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE Erg protein.
 GN ERG.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 OC Xenopodidae; Xenopus.
 OX NCBI_Taxid=8355;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE-20099678; PubMed-10633861;
 RA Baltzinger M., Mager-Heckel A.M., Remy P.;
 RT "Xlerg: Expression and pattern and overexpression during development
 RT plead for a role in endodermal cell differentiation.";
 RL Dev. Dyn. 216:420-433(1999).
 DR EMBL; AJ224126; CAB46567.1; -;
 DR HSSP; 001543; IFTL.
 DR InterPro: IPR000418; Ets.
 DR InterPro: IPR002341; HSF_ETS.
 DR InterPro: IPR003118; SAM_PNT.
 DR Pfam; PF00178; Ets; 1.
 DR Pfam; PF02198; SAM_PNT; 1.
 DR PRINTS; PR00454; ETSDOMAIN.
 DR SMART; SM00413; Ets; 1.

```
DR SMART; SM00251; SAM_PNT; 1.
DR PROSITE; PS00345; ETS_DOMAIN_1; 1.
DR PROSITE; PS00346; ETS_DOMAIN_2; 1.
DR PROSITE; PS50061; ETS_DOMAIN_3; 1.
SQ SEQUENCE 456 AA; 51482 MW; FFC137FD19099DED CRC64;
```

Query Match	77.98;	Score 2015.5;	DB 13;	Length 456;
Best Local Similarity	77.58;	Pred. No. 3.3e-149;		
Matches 383; Conservative	19;	Mismatches 23;	Indels 69;	Gaps 5;

[illegible][illegible]

QY	245	TTRPDLPEYQANRSMWTSHSPQSKATOPSSSTVPKTEDDQRPOLDYQILGTSRLAN	30
Db	225	PSKQDLSYPSRSMSTHNPAP - PKASQP - STVPKTEDDPRLDLPYQILGTSRLAN	28
QY	305	PGSGQIQLMQFLLELLSSSSNCLITWEGTNGEKKMDPDEVARRNGERKSKNNMYDGL	36
Db	283	PGSGQIQLMQFLLELLSSSSNCLITWEGTNGEKKMDPDEVARRNGERKSKNNMYDGL	34

QY	365	SRALRYDYKNTMTVHPHRESSMTKYRBDLPYMSYSGKRYAKAKFDPHGAGLADP----	42
Db	343	SRALRYDYKNTMTKY-----HGKRYAKKFPDPHGAGLADLPHPPE	38
QY	421	-----HAHPQKMTFVAPHPALPVTSSSFPAAPNPYMSPTGILPNR	46
Db	383	STWYKPSLPYMSYVAHPQKMTFVAPHPALPVTSSSFPAAPNPYMSPTGILPNR	44

QY	465	LPAAHPSHLGTY	478
Db	443	LPASHMSHLGTY	456
RESULT	8		
	099P061		

	Accession	Protein	Gene
AC	Q9P0U1	FABR1	401 AA
AD	Q9P0U1	FABR1	401 AA
DT	01-MAY-2000	(TREMBLrel. 13, Created)	
DT	01-MAY-2000	(TREMBLrel. 13, Last sequence update)	
DT	01-DEC-2001	(TREMBLrel. 19, Last annotation update)	
DE		Flil-1 protein.	
GN		Flil OR Flil-1.	
GN		Brachydanio rerio (zebrafish) (zebra danio)	

RA
RX
RN
RP
RQ
RR
RS
RT
RU
RV
RW
RX
RY
RZ
S
S1
S2
S3
S4
S5
S6
S7
S8
S9
S10
S11
S12
S13
S14
S15
S16
S17
S18
S19
S20
S21
S22
S23
S24
S25
S26
S27
S28
S29
S30
S31
S32
S33
S34
S35
S36
S37
S38
S39
S40
S41
S42
S43
S44
S45
S46
S47
S48
S49
S50
S51
S52
S53
S54
S55
S56
S57
S58
S59
S60
S61
S62
S63
S64
S65
S66
S67
S68
S69
S70
S71
S72
S73
S74
S75
S76
S77
S78
S79
S80
S81
S82
S83
S84
S85
S86
S87
S88
S89
S90
S91
S92
S93
S94
S95
S96
S97
S98
S99
S100
S101
S102
S103
S104
S105
S106
S107
S108
S109
S110
S111
S112
S113
S114
S115
S116
S117
S118
S119
S120
S121
S122
S123
S124
S125
S126
S127
S128
S129
S130
S131
S132
S133
S134
S135
S136
S137
S138
S139
S140
S141
S142
S143
S144
S145
S146
S147
S148
S149
S150
S151
S152
S153
S154
S155
S156
S157
S158
S159
S160
S161
S162
S163
S164
S165
S166
S167
S168
S169
S170
S171
S172
S173
S174
S175
S176
S177
S178
S179
S180
S181
S182
S183
S184
S185
S186
S187
S188
S189
S190
S191
S192
S193
S194
S195
S196
S197
S198
S199
S200
S201
S202
S203
S204
S205
S206
S207
S208
S209
S210
S211
S212
S213
S214
S215
S216
S217
S218
S219
S220
S221
S222
S223
S224
S225
S226
S227
S228
S229
S230
S231
S232
S233
S234
S235
S236
S237
S238
S239
S240
S241
S242
S243
S244
S245
S246
S247
S248
S249
S250
S251
S252
S253
S254
S255
S256
S257
S258
S259
S260
S261
S262
S263
S264
S265
S266
S267
S268
S269
S270
S271
S272
S273
S274
S275
S276
S277
S278
S279
S280
S281
S282
S283
S284
S285
S286
S287
S288
S289
S290
S291
S292
S293
S294
S295
S296
S297
S298
S299
S300
S301
S302
S303
S304
S305
S306
S307
S308
S309
S310
S311
S312
S313
S314
S315
S316
S317
S318
S319
S320
S321
S322
S323
S324
S325
S326
S327
S328
S329
S330
S331
S332
S333
S334
S335
S336
S337
S338
S339
S340
S341
S342
S343
S344
S345
S346
S347
S348
S349
S350
S351
S352
S353
S354
S355
S356
S357
S358
S359
S360
S361
S362
S363
S364
S365
S366
S367
S368
S369
S370
S371
S372
S373
S374
S375
S376
S377
S378
S379
S380
S381
S382
S383
S384
S385
S386
S387
S388
S389
S390
S391
S392
S393
S394
S395
S396
S397
S398
S399
S400
S401
S402
S403
S404
S405
S406
S407
S408
S409
S410
S411
S412
S413
S414
S415
S416
S417
S418
S419
S420
S421
S422
S423
S424
S425
S426
S427
S428
S429
S430
S431
S432
S433
S434
S435
S436
S437
S438
S439
S440
S441
S442
S443
S444
S445
S446
S447
S448
S449
S450
S451
S452
S453
S454
S455
S456
S457
S458
S459
S460
S461
S462
S463
S464
S465
S466
S467
S468
S469
S470
S471
S472
S473
S474
S475
S476
S477
S478
S479
S480
S481
S482
S483
S484
S485
S486
S487
S488
S489
S490
S491
S492
S493
S494
S495
S496
S497
S498
S499
S500
S501
S502
S503
S504
S505
S506
S507
S508
S509
S510
S511
S512
S513
S514
S515
S516
S517
S518
S519
S520
S521
S522
S523
S524
S525
S526
S527
S528
S529
S530
S531
S532
S533
S534
S535
S536
S537
S538
S539
S540
S541
S542
S543
S544
S545
S546
S547
S548
S549
S550
S551
S552
S553
S554
S555
S556
S557
S558
S559
S560
S561
S562
S563
S564
S565
S566
S567
S568
S569
S570
S571
S572
S573
S574
S575
S576
S577
S578
S579
S580
S581
S582
S583
S584
S585
S586
S587
S588
S589
S590
S591
S592
S593
S594
S595
S596
S597
S598
S599
S600
S601
S602
S603
S604
S605
S606
S607
S608
S609
S610
S611
S612
S613
S614
S615
S616
S617
S618
S619
S620
S621
S622
S623
S624
S625
S626
S627
S628
S629
S630
S631
S632
S633
S634
S635
S636
S637
S638
S639
S640
S641
S642
S643
S644
S645
S646
S647
S648
S649
S650
S651
S652
S653
S654
S655
S656
S657
S658
S659
S660
S661
S662
S663
S664
S665
S666
S667
S668
S669
S670
S671
S672
S673
S674
S675
S676
S677
S678
S679
S680
S681
S682
S683
S684
S685
S686
S687
S688
S689
S690
S691
S692
S69

RA Parental R.N., Shattuck A.V.;
 RA Insights into early vasculogenesis revealed by expression of the
 RA domain transcription factor Flt-1 in wild-type and mutant zebrafish
 R1 embryos.";
 RT Mech. Dev. 90:237-252(2000).
 RL EMBL: AJ246950; CAB56832.1.;
 DR HSSP: 001543; 1LFI.
 DR ZFIN: ZDR-GENE-980526-426. fl11
 DR

```

DR      InterPro; IPR000418; ETS.
DR      InterPro; IPR002341; HSF_ETTS.
DR      InterPro; IPR003118; SAM_PNT.
DR      Pfam; PF00178; ETS; 1.
DR      Pfam; PF02198; SAM_PNT; 1.
DR      PRINTS; PR00454; ETSDOMAIN.
DR      SMART; SMO0413; ETS; 1.
DR      SMART; SMO0251; SAM_PNT; 1.
DR      PROSITE; PS00345; ETS_DOMAIN_1; 1.
DR      PROSITE; PS00346; ETS_DOMAIN_2; 1.
DR      PROSITE; PS50061; ETS_DOMAIN_3; 1.
DR      SEQUENCE 451 AA; 50655 MW; F437D55D08E3549A CRC64;

```

Query Match	57.6%;	Score 1491.5;	DB 13;	Length 451;
Best Local Similarity	57.6%;	Pred. No. 2.7e-108;		
Matches 289;	Conservative 53;	Mismatches 85;	Indels 75;	Gaps 7

[illegible]

QY	120	TNERRYVAPDPPLMSTWDHAYROMLEMAVKEYGLPDPVDILLFONIDGKEJLCKMTRDDFOR	17
		
Db	118	TNERRYVAPDPSPMLSPDHYROMLEMAKEYGLEIDPIMFHSYDGEJLCKMKSDDFLR	17
		
QY	180	TPSYNADILLSHAHYRREPLPHLTSDDVDKALONSPMLHARNYGCATFIPFNVSYPE	23
		
Qb	178	TSYVNEFVLLSHYNTYRRES	20
	SSISYVNTFSSHAD	

QY 240 ATGRTTTRDPLPEAKRRKAMTSHHPQSKATQDSSSTYPKTEDAPQLDPIQLGPPS 29
Db 210 QSPRLAKDASTDAVRRKGGNNHSHSGSGSPYVSSQSKNPQPPQDPPQLIGPPS 26
QY 300 SRLPAPSGQIQLOMDFLELLSDSSNSNCITWETGNEBFKMTDPDEVARRRGERSKSPM 35
Db 270 SRLPAPSGQIQLOMDFLELLSDSSNSNACITWETGNGEFTKMTDPEVARRRGERSKSPM 32

QY	360	NYDLNLALRYYYDKIMIKKYNPPSSMKYTPSDLPMWSHGCRARYAKDFHGLIALO	41
Db	330	NYDLKSLALRYYYDKIMIKY-----HGRRAYKKDFGLIALO	36
QY	420	P-----HAHOKMNFVALPHPALPYTSSSEFAAPRPYNSSPTGCI	45
Db	370	PHPLPSMKYKPELPLVYSYAHOOOKNFVSPHPSMSPYTNSSNFEPPTTPYSSSPGCI	42

Oy	460	YPNTRLPA--AHMPSHLCTYY	478
		: : : : : :	
Db	430	YPNPSVPRHANSHPSHGLSXY	451

ID	PRELIMINARY	PRJ	432 AA.
053425			
093425			
01-NOV-1998	(TREMBLrel. 08, Created)		
01-NOV-1998	(TREMBLrel. 08, Last sequence update)		
01-JUN-2001	(TREMBLrel. 17, Last annotation update)		
DE	FLI transcription factor.		
GN	FLI.		
GN	Coturnix coturnix tanocica (Japanese quail)		

OC Actinosauria; Aves; Neognathae; Galliformes; Phasianinae; Phasianinae
OC Columbidae;
OX NCBI_TaxID=93934;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98358003; PubMed=9694627;
RA Mager A.M., Graplin-Botton A., Ladjali K., Meyer D., Wolff C.M.,

RT "The avian fil gene is specifically expressed during embryogenesis in
RT a subset of neural crest cells giving rise to mesenchyme.";
RL Int. J. Dev. Biol. 42:561-572(1998).
DR EMBL: Y14773; CAA75077.1; -;
DR EMBL: Y14774; CAA75078.1; -;
DR HSSP: Q01543; IPII.
DR InterPro: IPR000418; Ets.
DR InterPro: IPR002341; HSF_ETTS.
DR InterPro: IPR003118; SAM_PNT.
DR Pfam: PF00178; Ets; 1.
DR Pfam: PF02198; SAM_PNT; 1.
DR PRINTS: PRO0454; ETSDOMAIN.
DR SMART: SM00413; ETS; 1.
DR SMART: SM00251; SAM_PNT; 1.
DR PROSITE: PS00345; ETS_DOMAIN_1; 1.
DR PROSITE: PS00346; ETS_DOMAIN_2; 1.
DR PROSITE: PSS0061; ETS_DOMAIN_3; 1.
SQ SEQUENCE 432 AA: 48761 MW: 83259668CB104EB CRC64;

Query Match	55.3%	Score 1432;	DB 13	Length 432;
Best Local Similarity	57.1%;	Pred. No. 1.1e-103;		
Matches 287; Conservative	48;	Mismatches 72;	Indels 96;	Gaps 10

[illegible]

RESULT 10		
09PUL6		
ID	09PUL6	PRELIMINARY; PRT; 414 AA.
AC	09PUL6;	
DT	01-MAY-2000 (TREMBLrel. 13, Created)	
DT	01-MAY-2000 (TREMBLrel. 13, last sequence update)	
DT	01-DEC-2001 (TREMBLrel. 19, last annotation update)	
DE	F111 protein (fragment) .	
GN	F111.	
OS	Brachydanio rerio (Zebrafish) (Zebra danio) .	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes	
CC	Cyprinidae; Danio.	

OX NCBI_TaxID=7955;
RN
RP
SEQUENCE FROM N.A.
RX MEDLINE=98294174; PubMed=9630750;
RA Thompson M.A., Ransom D.G., Pratt S.J., Maclean H., Kieran M.W.,
RA Derrick H.W.I.I., Vail B., Huber T.L., Paw B., Brownlie A.J.,
RA Oates A.C., Flicz A., Gates M.A., Amores A., Bahary N., Talbot W.S.,
RA Her H., Beier D.R., Postlethwait J.H., Zon L.I.;
RT "The cloche and spadetail genes differentially affect hematopoiesis
and vasculogenesis ";
RL Dev. Biol. 197;248-269(1998).

Query Match	51.8%	Score 1341	DB 13	Length 414
Best Local Similarity	55.9%	Pred. No. 1.4e-96		
Matches 259, Conservative	50	Mismatches 82	Indels 72	Gaps 6

QY	14	EDOSIFECAY-GSHPLKTEMTASSEYEGOTSMKSRVQOOLSOPEARVITIKMCNP	72
Db	1	EDOSLEPPRYAANAPLPEKTDMTASGTODYQGTHTKINLIPPOQEING-PYRVANKREY--	57
QY	73	NOVNGSRNSPDSCVAKGKMKVSSSDNVGNANTYSYMEKHIPPNMTTNERVIYPADPT	1322
Db	58	DHNGSRSPDVCDSVGKCNKKVGGTEGASOMNYTGYMDEKCAPPPNMTTNERVIYPADPS	1171
QY	133	LMSTDHVRONTLEMAVKEGLDPDVOILFONINDGKELCMKTKRDEPQRLTPTSYNDIILSHL	1922
Db	118	LMSTDHVRQKMDMAIKKEYGLQEDITLAFHSTIDGELCKMSKDFLRLTFSYNTVEVLISHL	1777
QY	193	HYLEETPLPHLTSDVDVKALONSPRLMHNARNTGATEIFPNTSYVEPATORITTRDPDLPY	2522
Db	178	NYLRES-----SSSIYNTPSHADQSPRLAKADASY	2099
QY	253	EQARRSAMTSHSHPLOSKATQPOSSSYTPKTEEDARPOLDPOVQILGPTSSRLANQSGOIOL	3122
Db	210	DAVREITGMSNNHGKSPVIVGSOVSKNPDPQPODPOVQILGPTSSRLANQSGOIOL	2699
QY	313	WQFLLELSDSNINCITTEWGTNEEFMTDPDEVARERGERKSKPNMYDKLSRLALRYYY	3722
Db	270	WQFLLELSDSANAGCITTEWGTNEEFMTDPDEVARERGERKSKPSNMYDKLSRLALRYYY	3399
QY	373	DKNIMTVHPPESSMYKYPSDLPYMSSYHGKRYAYKDFEHGIAOALOP-----	4202
Db	330	DKTMTKY-----HKRYAYKIDFEGIAOALOPHPTESMTWKYPS	3699
QY	421	-----HAHPKMNVPARHPALPVTTSSSFFAAPRYWNSP	455
Db	370	ELPVPVSTHAHQQVKNVNSHPBPMPATVTSNNFGPRTTPYVSSP	412

RESULT 11
 Q9BZD1 PRELIMINARY; PRT; 476 AA.
 AC Q9BZD1
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE Ewings sarcoma EWS-Flil (type 1) oncogene.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NCBI_TaxID=9606;
 RX [1]
 RP SEQUENCE FROM N.A.
 RA Uhl T.B.S., Rainey D.R., Craig A.R., Meredith D.M., Lewis I.J.;
 RT "The Ewings sarcoma EWS-Flil (type 1) oncogene."
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF327066; AKL1227.1; -
 DR HSPR: Q01543; FLI1
 DR InterPro: IPR000418; Ets.
 DR InterPro: IPR002341; HSF_ETS.
 DR Pfam: PF00178; Ets; 1.
 DR PRINTS: PRO0454; ETSDOMAIN.
 DR SMART: SM00413; ETS; 1.
 DR PROSITE: PS00345; ETS_DOMAIN_1; 1.
 DR PROSITE: PS00346; ETS_DOMAIN_2; 1.
 DR PROSITE: PS50061; ETS_DOMAIN_3; 1.
 SQ SEQUENCE 476 AA; 51917 MW; AA76A1B6CAE7A90 CRC64;

Query Match 32.1%; Score 829.5; DB 4; Length 476;
 Best Local Similarity 40.9%; Pred. NO. 1.5e-56;
 Matches 202; Conservative 46; Mismatches 147; Indels 99; Gaps 16;
 QY 23 YGSPHLAKTEMT---ASSSEYQGTSKMSPRVPOQDWLSQPARVITIM--ECNPNQVNG 77
 DB 44 YGQP---TGVSYQAQGTATYTGTAATAT---YGPRTGTYTTTAQAASQAPQG 92
 QY 78 SRNPPDCSVA---KGGKVVSSSDVNGVNYGMEKHITPPNMTNERRVIVPADPT 132
 DB 93 YGTGAYPTTATVTTTQASYAASAYGTQAPAYGQOPATAPTRPDGKPTETSQFQ 152
 QY 133 LMSVDHVRQWLEMAVKEGLPDVDLLEFQNDGKELCKMTDDFORLT--PSYADILLS 190
 DB 153 SSTGGVYQPSLGGYQOSNYSYQVP---GSPYQPVYAPPSY-----P 191
 QY 191 HLHLETRPLPHTLSDVDKALQNSPRLMARNTGATFIF---PNTSVYEATQRIITR 247
 DB 192 PTSSSQPTSTYQSSSQNTYQPSYQSSSYGQSSYGGQPPTS--YPTQGSTSQ 250
 QY 248 PDLPYEQARRSAMWTSHPPTQSKATOPSSSTVPKTEQDPQLADPYQILGPTSSRLANPGS 307
 DB 251 PS-QYSQ-QSSSYGQSPPL-----GGAQTSKNTEDQRPDPYQILGPTSSRLANPGS 302
 QY 308 GQIQLMQFLLELSDSSNCTITWEGTNGEFGKMTDPDEVARRWGERKSKPMANTDKLSRA 367
 DB 303 GQIQLMQFLLELSDSSNCTITWEGTNGEFGKMTDPDEVARRWGERKSKPMANTDKLSRA 362
 QY 368 LRYYYDKNIMTKVHPRESSMTKYPSPDLRYMSSYGRKRAYFDEHGAQALQP----- 420
 DB 363 LRYYYDKNIMTKV-----HGKRAYFDEHGAQALQPPTSSSM 402
 QY 421 -----HAPQKMFVAPHPALPVTSSTSSFFAAPNPNWNSPTGTGTYPN---TR 464
 DB 403 YKYPSSDISYMPHYAHOQKVFVPPHPSMPVTSSSFFGAASQWTSPTGTGTYPNPNVR 462
 QY 465 LPAHMPSHLGTYY 478
 DB 463 HPNTHPSHLGSYY 476

RESULT 12
 Q95T62 PRELIMINARY; PRT; 257 AA.
 ID Q95T62

AC Q95T62;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE GH18452P.
 GN EWS65A OR CG7018.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NCBI_TaxID=7227;
 RX [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN-BERKELEY;
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
 RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
 RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
 RA Nunoo J., Paclad J., Parragas V., Park S., Phouanavong S., Wan K.,
 RA Yu C., Lewis S.E., Rubin G.M., Gelinker S.;
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY060316; AAL2535.1; -
 DR FlyBase: FBgn0005658; Ets65A.
 DR InterPro: IPR000418; Ets.
 DR InterPro: IPR002341; HSF_ETS.
 DR Pfam: PF00178; Ets; 1.
 DR PROSITE: PS00345; ETS_DOMAIN_1; UNKNOWN_1.
 DR PROSITE: PS00346; ETS_DOMAIN_2; UNKNOWN_1.
 DR PROSITE: PS50061; ETS_DOMAIN_3; 1.
 SQ SEQUENCE 257 AA; 28700 MW; 4C6319AF07E5DAB0 CRC64;

Query Match 23.3%; Score 602.5; DB 5; Length 257;
 Best Local Similarity 49.1%; Pred. NO. 3.3e-39;
 Matches 139; Conservative 31; Mismatches 46; Indels 67; Gaps 12;
 QY 221 ARNNGAFTFPNTSVYEATQRIITRDLPEQARRSAMWTSHPPTQSKATOPSSSTVP 280
 DB 12 SRSTGS---FSNFSM-----LPADSSYKSSWGSHS-STQSGY--SSNAG 51
 QY 281 KTEQORPOL---DPYQILGPTSSRLANPGSGQIQLMQFLLELSDSSNCTITWEGTNGE 337
 DB 52 IKDHPHSLQRPDPYQMGPTSSRLASSGQIQLMQFLLELSDSSNCTITWEGTNGE 111
 QY 338 FKMTDPDEVARRWGERKSKPMANTDKLSRALRYYYDKNIMTKVHPRESSMTKYPSPDLPYM 397
 DB 112 FKMTDPDEVARRWGERKSKPMANTDKLSRALRYYYDKNIMTKV----- 154
 QY 398 SSYGKRAYFDEHGAQALQPNH-----HPQK-MFYAPHPALPV 439
 DB 155 ---HGKRAYFDEHGAQALQPNH-----HPQK-MFYAPHPALPV 439
 QY 440 TSSSFF--AAPNPNWNSPTGTGTY--PN--TRLPAHMPSHLGTYY 477
 DB 211 SSASIFPSAASWGMGSPATNLYOPHSMSHYTPSHVAPHLSSY 253

RESULT 13
 Q95ZEO PRELIMINARY; PRT; 164 AA.
 ID Q95ZEO
 AC Q95ZEO;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE Ery protein (Fragment).
 GN Ery.
 OS Hediste diversicolor.
 OC Eukaryota; Metazoa; Annelida; Polychaeta; Palpata; Aciculata;
 OC Phyllodoidea; Hediste.
 NCBI_TaxID=126592;
 RX [1]
 RP SEQUENCE FROM N.A.
 RA Bocquet B.;
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.

```

RN      SEQUENCE FROM N.A.  
RP  
RC STRAIN-DIVERSICOLOR;  
RA Bocquet-Michembled B.;  
RL Thesis (2000), Department of Biologie Animale, Université de Lille,  
RR Villeneuve d'Ascq, France.  
DR EMBL; AJ311813; CAC44038.1; -  
DR InterPro; IPR000418; ETS.  
DR InterPro; IPR002341; HSF_ETSS.  
DR Pfam; PF00178; Ets_1.  
DR PROSITE; PS00346; ETS_DOMAIN_1; UNKNOWN_1.  
DR PROSITE; PS00345; ETS_DOMAIN_2; UNKNOWN_1.  
DR PROSITE; PSS0061; ETS_DOMAIN_3; 1.  
FT NON_TER          1         1  
SQ     SEQUENCE    164 AA;   164       MM;  78SA3897643CDBA9 CRC64;  
  
Query Match           20.1%; Score 519.5; DB 5; Length 164;  
Best Local Similarity 58.2%; Pred. No. 5.3e-33;  
Matches 103; Conservative 14; Mismatches 21; Indels 39; Gaps 3.  
  
QY      DPLYGLGTFSTSLANPGSGQLDLMFLELDLSDSNSNCITEGTNGEFGKMTDPDEVARR 349  
|||: | ||||:|||||:|||||:|||||:|||||:|||||:|||||:  
Db      9 DPYLGFGISRSLHSISGGQIQDMOFLFILSDSANSHITTMEGTNGEKVLDPDENVR 68  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
QY      350 WGERSKSPNMNYDKLSRALRYYYDNKITKYHPPESSMYKKPSDLPYMSSYGKRARYKF 409  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
Db      69 WGERSKSPNMNYDKLSRALRYYYDNKITKV-----HGKRAYKF 108  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
QY      410 DEHGIQAOLQP-----HAHQKMNFVAPPPALPVYSSEFPAR 448  
|||:|||:|||::|||::|||::|||::|||::|||::|||:  
Db      109 DPAGIAQMOPSTTPPAAYKTQQDILMSGYHTSNTLNIGAHAP-MASSASGEFFPP 164  
  
RESULT 14  
ID Q22355 PRELIMINARY; PRT: 377 AA.  
AC Q22355.  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-JAN-1999 (TREMBLrel. 09, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE Hypothetical 41.5 kDa protein.  
GN T0B84.3.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdiltida; Rhabditoidea;  
CC Rhabdilidae; Peloderinae; Caenorbablitia.  
OX NCBI_Taxid=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL NZ;  
RX MEDLINE=99069613; PubMed=9851916;  
RA None;  
RT "Genome sequence of the nematode C. elegans: A platform for investigating biology. The C. Elegans Sequencing Consortium."; Science 282:2012-2018(1998).  
RL [2]  
RN RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL NZ;  
RA Stellyes L.;  
RT "The sequence of C. elegans cosmid T0B84.";  
RL Submitted (NOV-1995) to the EMBL/Genbank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL NZ;  
RA Waterston K.;  
RT "Direct Submission."  
RL Submited (SEP-2001) to the EMBL/Genbank/DBJ databases.  
DR EMBL; U39470; AAC71120.1; -.  
DR HSSP; O01543; IFLI.  
DR InterPro; IPR000418; Ets.  
DR InterPro; IPR002341; HSF_ETSS.  
DR Pfam; PF00178; Ets_1.  
DR PRINTS; PR00454; ETSDOMAIN.
```

[illegible]

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 9, 2002, 16:55:42 ; Search time 3095.42 seconds
(without alignments)
4494.100 Million cell updates/sec

Title: us-09-902-772-4
Perfect score: 2588
Sequence: 1 MASTKEALSVSEDSLE.....IYPNTRLPAAHMPHSLGTY 478

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 segs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+ p2n.model -DEV=xlp
-O/cpnt2.1/USPTO/US90902772/rnat.07112002.134033.9031/app.query.fasta_1.1294
-Db=GenEmbl -OPMT-fastext -SUFFIX=rge -MINMATCH=0.1 -IOPTCL=0 -DELOPXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pcr -NOR-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USBR=US90902772.@CN1.1.3441.@rnat.07112002.134033.9031 -NCPU=6 -ICPU=3
-NO_XLPHY -NO_MMAP -LARGEIOERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120
-MARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl.*
1: gb.ba:*
2: gb.htg:*
3: gb.in:*
4: gb.om:*
5: gb.ov:*
6: gb.pat:*
7: gb.ph:*
8: gb.pl:*
9: gb.pr:*
10: gb.ro:*
11: gb.sts:*
12: gb.sy:*
13: gb.un:*
14: gb.vl:*
15: em.ba:*
16: em.fun:*
17: em.in:*
18: em.hum:*
19: em.mu:*
20: em.om:*
21: em.or:*
22: em.ov:*
23: em.pat:*
24: em.ph:*
25: em.pl:*
26: em.ro:*
27: em.sts:*
28: em.un:*

29: em.vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_man:*
37: em_htg_vtl:*
38: em_sy:*
39: em_higo_hum:*
40: em_higo_mus:*
41: em_higo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2436	94.1	1516	5 CGERG	X77159 G.gallus ER
2	2436	94.1	1528	6 E31254	E31254 Protein hav
3	2293.5	88.6	2209	10 AB073080	AB073080 Mus muscu
4	2262.5	87.4	1413	5 AY065661	AY065661 Gallus ga
5	2262.5	87.4	1447	6 E31253	E31253 Protein hav
6	2186	84.5	2133	10 AB073079	AB073079 Mus muscu
7	2180.5	84.3	3166	9 HUMERG2	M17254 Human erg2
8	2180	84.2	1798	5 XLAJ4125	AJ224125 Xenopus 1
9	2177.5	84.1	1808	10 AB031088	AB031088 Rattus no
10	2149.5	83.1	2131	10 AB073078	AB073078 Mus muscu
11	2015.5	77.9	2012	5 XLAJ4126	AJ224126 Xenopus 1
12	1808.5	69.9	3126	9 HUMERG11	M21535 Human erg p
13	1491.5	57.6	1436	5 DRE249580	AJ249590 Danio rer
14	1482	57.3	1359	9 AY029368	AY029368 Homo sapi
15	1482	57.3	2916	6 BC010115	BC010115 Homo sapi
16	1482	57.3	2916	6 A36461	A36461 Sequence 2
17	1482	57.3	2938	6 AR080101	AR080101 Sequence
18	1482	57.3	2938	9 HSHMFL1	X67001 H.sapiens H
19	1482	57.3	2957	9 HUMERGBFL1	M98833 Homo sapien
20	1482	57.3	2959	9 BC001670	BC001670 Homo sapi
21	1482	57.3	6849	12 AY029367	AY029367 Synthetic
22	1479	57.1	1673	9 S45205	S45205 FLI-1-Frien
23	1474	57.0	1729	10 MMFL1	X59421 Mouse FLI-1
24	1440.5	55.7	4403	5 XLFILG	X66979 X.laevis mR
25	1432	55.3	3490	5 CCNNAFL1	Y14773 Coturnix co
26	1420	54.9	3545	5 CCEFLIONCO	Y14774 Coturnix co
27	1394	53.9	1932	9 HUMFL1A	M93255 Human FLI-1
28	1341	51.8	1245	5 AF177538	AF177538 Danio rer
29	1253.5	48.4	816	10 S66169	S66169 Erg-3-immun
30	1036.5	40.1	762	9 S72621	S72621 EMS...erg 1
31	834	32.2	1431	9 AF327066	AF327066 Homo sapi
32	809.5	31.3	14518	9 AP001423	AP001423 Homo sapi
33	809.5	31.3	100000	9 AP000021	AP000021 Homo sapi
34	809.5	31.3	100000	9 AP000163	AP000163 Homo sapi
35	809.5	31.3	340000	9 AP001731	AP001731 Homo sapi
36	778	30.1	588	9 S72865	S72865 EMS...EMS-e
37	702.5	27.1	66352	9 HSI17293	Y17293 Homo sapien
38	702.5	27.1	187131	2 AP001535	AP001535 Homo sapi
39	702.5	27.1	217382	9 AP001122	AP001122 Homo sapi
40	604	23.3	2180	3 AY060316	AY060316 Drosophi
41	602	23.3	634	3 SULERG	M81067 Sea urchin
42	514.5	19.9	491	3 HDJ11813	AJ311813 Hediste d
43	485.5	18.8	169741	9 AC097468	AC097468 Homo sapi
44	484.5	18.7	83830	2 AC095777	AC095777 Rattus no
45	484.5	18.7	192968	2 AC127107	AC127107 Rattus no

RESULT 1

ALIGNMENTS

LOCUS GGERG 1516 bp mRNA linear VRT 27-APR-1995
DEFINITION G.gallus ERG mRNA.
ACCESSION X77159
VERSION X77159.1 GI:790439
KEYWORDS ERG gene.
SOURCE chicken.
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 1516)
Dhotadain, P., Dewitte, F., Desbiens, X., Stehelin, D. and
Duterque-Coquillard, M.
TITLE Mesodermal expression of the chicken erg gene associated with
precartilaginous condensation and cartilage differentiation
JOURNAL Mech. Dev. 50 (1), 17-28 (1995)
MEDLINE 95329425
PubMed 7605748
REFERENCE 2 (bases 1 to 1516)
Duterque-Coquillard, M.
AUTHORS Direct Submission
TITLE Submitted (17-JAN-1994) M. Duterque-Coquillard, CNRS UA 1160,
JOURNAL Oncologie Molculaire, Institut Pasteur, 1 rue Calmette, 59019
Lille, FRANCE

FEATURES
source Location/Qualifiers
1..1516
/organism="Gallus gallus"
/db_xref="taxon:9031"
/tissue_type="spleen"
/clone_lib="lambda gcl10"
/dev_stage="adult"
63..1499
/gene="ERG"
63..1499
/gene="ERG"
/codon_start=1
/protein_id="CA54404.1"
/db_xref="GI:790440"
/translation="MASTIKELSVSSEDOSLFECAYSGPHLAKTEMASSSEYGT
SKMSPRVPOODMLSQPARVTIKMECNONVNSNDSDSVAKGKMSVDVGM
NYGSMERKHIPPMNTNERBIVYPAPITLMSDHYVQMLEAKKEIGLVDVLLF
ONIDKEIKKTKDFOPLCTSYNDILSLHUTRETPPLHLISDDVKALQNSPL
MHARNTGATGIFPNTSVYPEATORTITRPDLPYEQARSAMTSHSHTQSKATQPS
STVPTEDORQDLPYQILGPTSSRLAPSGGQIQLOMFLLELSDSSNSNCITWEGT
NGEEMTDPEVARMEGRKSKPMKNYDKLSRALRYYYDKNIMTVHGKRAYKDFH
GIAOLQHPRESSMYKXSDLPYSSSYHAHPQKNFYAHPHPALPVTSSEFAAPNP
YWSPTGGIYENTRLPAHMPHSICTTY"

BASE COUNT 458 a 392 c 327 g 339 t
ORIGIN

Alignment Scores:
Pred. No.: 2,466-158 Length: 1516
Score: 2436.00 Matches: 458
Percent Similarity: 91.97% Conservative: 0
Best Local Similarity: 91.97% Mismatches: 0
Query Match: 94.13% Indels: 40
DB: 5 Gaps: 2

US-09-902-772-4 (1-478) x GGERG (1-1516)

QY 1 MetAlaSerThrIleLysGluAlaLeuSerValValSerGluAspGlnSerLeuPheGlu 20
DB 63 ATGGCAGACGATTAAGAGACATTATCATGCGTAGAGAGACAGACGCTGTTTGG 122

QY 21 CysAlaTyrGlySerProHisLeuAlaLysThrGluMetThrAlaSerSerSerGlu 40
DB 123 TGTGCTACGATGCGCCACCTTCGCAAGACAGAAATACAGACCTCCCTCCAGTGAA 182

QY 41 TyrGlyGlnThrSerLysMetSerProArgValProGlnGlnAspTyrLeuSerGlnPro 60
DB 183 TATGGCAAAACATCAAAAGTAGCGCGCGCTTCCCGACAGAGACTGGTTATCACAGCCC 242

QY 61 ProAlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsn 80
DB 243 CCGGCCAGAGTACCATTAAGATGAGGTAAACCCAAACAGGTTAATGGTCAAGAAAT 302

QY 81 SerProAspAspCysSerValAlaLysGlyLysMetValSerSerSerAsnVal 100
DB 303 TCACCTGATGATGCGACCGCTGGCAAAAGAGGAAATAGTTAGCATGTCACACATGTT 362

QY 101 GlyMetAsnTyrGlySerTyrMetGluGluLysHisIleProProAsnMetThrThr 120
DB 363 GGGATGAACATTAGGAAGCTACATGGAAGAAGACATATTCCCTCCCAATATGACAAAC 422

QY 121 AsnGluArgValIleValProAlaAspProThrLeuTyrSerThrAspHisValArg 140
DB 423 AATGAACGAAGATTATTGTGCAGAGATCTACGTTATGAGACAGACCATGTACGG 482

QY 141 GlnTyrPheGlnTyrPalaValLysGluTyrGlyLeuProAspValAspIleLeuLeuPhe 160
DB 483 CAGTGGCTGGAGTGGGCACTGAAGAGATGCTTCCTCCAGAGTGAGACATCTTGTTGCTC 542

QY 161 GlnAsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArgLeuThr 180
DB 543 CAGAACATTGATGGCAAGAGATTGTGTAATGACCAAGATGACTTCCAGAGACTCAGC 602

QY 181 ProSerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGluThrProLeu 200
DB 603 CCGAGTATAACGACGATATCTCTGTCACACTACACTACTCAGAGAGACTCTCTT 662

QY 201 ProHisLeuThrSerAspAspValAspLysAlaLeuGlnAsnSerProArgLeuMetHis 220
DB 663 CCACATTGACTTCAGATGATGATGTATAGGCTTACAAACCTCCACCGGTTAATGCAT 722

QY 221 AlaArgAsnThrGlyGlyAlaThrPheIlePheProAsnThrSerValTyrProGluAla 240
DB 723 GCTAGAAACACAGAGAGACCCACTTTTATTTTCCAAATACATCAGTTTACCAGAGCA 782

QY 241 ThrGlnArgIleThrThrArgProAspLeuProTyrGluGlnAlaArgSerAlaTyrP 260
DB 783 ACGCAAAAGAAATAAACAAGGCCAGATTAACTTATGACCAAGCAGAGATCAGCGTGG 842

QY 261 ThrSerHisSerHisProThrGlnSerLysAlaThrGlnProSerSerSerThrValPro 280
DB 843 ACGAGTCACAGCCATCCCTCAGTCACAAAGCTACCAACATCTTCAACACTGCGCC 902

QY 281 LysThrGlnAspGlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSerSer 300
DB 903 AAAACAGAAGACACAGCTCCTCAGTTAGATCTTATCAGATTCTTGACCGACCGACAC 962

QY 301 ArgLeuAlaAsnProGlySerGlyGlnIleGlnLeuTyrPglPheLeuLeuGluLeuLeu 320
DB 963 CGCTTGCAAATCCAGAGGATGGGACAGATACAGTATGCGACTTCTCAGAGCTTCTG 1022

QY 321 SerAspSerSerAsnSerAsnLysIleThrTyrPglGluGlyThrAsnGlyLysMet 340
DB 1023 TCGGACAGCTCCAACTCCAAAGCATGCACTCGGAGGCGCCAAAGGGAGATTCAAGATG 1082

QY 341 ThrAspProAspGluValAlaArgArgTyrPglGlyGluArgLysSerLysProAsnMetAsn 360
DB 1083 ACAGACCTCATGATGAGTGGCTGGCTGGGAGAGAGAGAAAGCAAACTTACATGTAAC 1142

QY 361 TyrAspLysLeuSerArgAlaLeuArgTyrTyrTyrAspLysAsnIleMetThrLysVal 380
DB 1143 TATGACAAACCTAGCGCGTACTTGCATCTATGACAAATAATATATATGACTAAATG 1202

QY 380 ----- 380

DB 1203 CATGTAAACGCTATGCTTACAAATTTGATTTCCAGGAAATGCTCAGGCTCCAGCCT 1262

QY 381 HisProGlnGluSerSerMetTyrLysTyrProAspAspLeuProTyrMetSerSerTyr 400
DB 1263 CACCTCCAGATCATCTCATGTACAAATATCCCATGAGACTCCCTTACATGAGTTCTTAC 1322

QY 401 HiscllylsarqlyrAlaTyrLysPheaspheHiscllylleAlaGlnAlaLeuGlnPro 420
Db 1322 ----- 1322
QY 421 HisAlaHisProGlnLysMetAsnPheValAlaProHisProAlaLeuProValThr 440
Db 1323 CATCACACCCCGAAGATGAACTTGTAGCTCCCATCCCTGGCTTGGCCGTAACC 1382
QY 441 SerSerSerPhePheAlaAlaProAsnProTyrTrpAsnSerProThrGlyGlyLeTyr 460
Db 1383 TCATCCAGCCTTTTGTGTCGCCCTAATCCATCTGGAATTCACCACTGGAGGCACTTAC 1442
QY 461 ProAsnThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyr 478
Db 1443 CCCAATACGAGCTGCAGCTGCTCATATGCTTCCATCTTGGCACCCTACTAC 1496
RESULT 2
LOCUS E31254 1528 bp DNA linear PAT 18-JUN-2001
DEFINITION Protein having cell calcifying inhibitory activity and gene
encoding the same.
ACCESSION E31254
VERSION E31254.1 GI:13025686
KEYWORDS JP 199075871-A/2.
SOURCE unidentifed.
ORGANISM unidentifed.
REFERENCE 1 (bases 1 to 1528)
AUTHORS Hiroyasu,I., Yoshinobu,H., Marjio,P., Joel,R. and Helena,E.
TITLE Protein having cell calcifying inhibitory activity and gene
encoding the same
JOURNAL Patent: JP 199075871-A 2 23-MAR-1999;
CHUGAI PHARMACEUT CO LTD, HE TRUSTEES OF THE UNIVERSITY OF
PENNSYLVANIA
COMMENT OS Unidentifed
PN JP 199075871-A/2
PD 23-MAR-1999 JP 1998166076
PF 29-MAY-1998 JP 1998166076
PR 18-JUN-1997 US 08/878177,20-JUN-1997 US 60/050297 PI
HIOKASU IWAMOTO,YOSHINOBU HIGUCHI,MARJIO PASHIRIKI, PI JOEL
ROZENBURGOM.
PI HELENA E
PC C12N15/09,A61K48/00,C07K14/465,C07K16/18,C12Q1/68//A61K38/00,
PC C12P21/02,
PC C12P21/08,C12N15/00,A61K37/02
CC Strandedness: Single;
CC Topology: Linear;
PH key Location/Qualifiers
FT source 1..1528
FT 1..1528
FEATURES
source Location/Qualifiers
1..1528
/organism="unidentifed"
BASE COUNT 464 a 395 c 330 g 339 t
ORIGIN
Alignment Scores:
Pred. No.: 2.48e-158 Length: 1528
Score: 2436.00 Matches: 458
Percent Similarity: 91.97% Conservative: 0
Best Local Similarity: 91.97% Mismatches: 0
Query Match: 94.13% Indels: 40
DB: 6 Gaps: 2
US-09-902-772-4 (1-478) x E31254 (1-1528)
QY 1 MetalaserThrIleLysGlnAlaLeuSerValValSerGluAspGlnSerLeuPheGln 20
Db 63 ATGCAAGCACTATTAGGAAGCATATCAAGTGTGAGTGAAGCAAGCAAGCTCTTGGAG 122
QY 21 CysAlaTyrGlySerProHisLeuAlaLysThrGluMetThrAlaSerSerSerGlu 40
|||||

Db 123 TGTGCTTAAGGATGCCCCACCTTGCAAAGACACAAATGACAGCTCTCTTCCAGTGA 182
QY 41 TyrGlyGlnThrSerLysMetSerProArgValProGlnGlnAspTrpLeuSerGlnPro 60
Db 183 TATGGCAACATCAAAAGATGAGCCCGCGGCTTCCAGAGAGACTGGTTATCACAGCC 242
QY 61 ProAlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsn 80
Db 243 CCGGCCAGAGTACCATTAAAGATGAGTGAACCCAAACAGGTTAATGGCTCAAGAAAT 302
QY 81 SerProAspAspCysSerValAlaLysGlyGlyLysMetValSerSerSerAspAsnVal 100
Db 303 TCACCTGATGACTGACGCGTGGCAAAAGAGGAAATGGTTACAGTTCAGACAAATTT 362
QY 101 GlyMetAsnTyrGlySerTyrMetGluGluLysHisIleProProAsnMetThrThr 120
Db 363 GGGATGAACATATGAGAGCTATGAGAGAGAGATNTCCGCCCTCCAAATATGACAAC 422
QY 121 AsnGluArgArgValIleValProAlaAspProThrLeuTrpSerThrAspHisValArg 140
Db 423 AATGACGAGAGTTAATGTTGTCAGAGATCTTACGTTATGAGACACAGACCATGTACGG 482
QY 141 GlnThrLeuGluTrpAlaValLysGluTyrGlyLeuProAspValAspIleLeuPhe 160
Db 483 CAGTGGCTGAGTGGCGAGTGAAGAGATATGCTTCCAGAGTGGACATCTTGTTC 542
QY 161 GlnAsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArgLeuThr 180
Db 543 CAGAAACATGATGGGAAGAGTGTGTAAATGACAAAGATGACTTCCAGAGACTCAG 602
QY 181 ProSerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGluThrProLeu 200
Db 603 CCGAGCTAATACGAGATATCTCTCTGACACACTACTACCTCAGAGAACTCTCTTT 662
QY 201 ProHisLeuThrSerAspAspValAspLysAlaLeuGlnAsnSerProArgLeuMetHis 220
Db 663 CCACATTTGACTTGATGATGATGATTAAGGCTTACAAACTCTCCAGGTTAATGAT 722
QY 221 AlaArgAsnThrGlyGlyAlaThrPheIlePheProAsnThrSerValTyrProGluAla 240
Db 723 GCTGAAGAACAGAGAGAGAGCACTTATTTTCCAAATATCATCAGTTACCAGAAACA 782
QY 241 ThrGlnArgIleThrThrArgProAspLeuProTyrGluGlnAlaArgArgSerAlaTrp 260
Db 783 ACGGAAGAAATAAACAACAAGGCGAGATTACCTATGAGCAAGGAGAGATCAGCGTG 842
QY 261 ThrSerHisSerHisProThrGlnSerLysAlaThrGlnProSerSerSerThrValPro 280
Db 843 ACGAGTCACAGCCATCCCATCAGTCAAAAGCTTACCACCATCATCTTCAACAGTGGCC 902
QY 281 LysThrGluAspGlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSerSer 300
Db 903 AAAACAGAAAGCAAGCGCTCAGTATGATCTTATCGATTTCTTGAGCCAGCAGCGC 962
QY 301 ArgLeuAlaAsnProGlySerGlyGlnIleGlnLeuTrpGlnPheLeuLeuLeuLeu 320
Db 963 CGTCTGCAATCCAGGAGAGTGGCAGATACAGTATGAGCAGTTCCTACTGGAGCTTCTG 1022
QY 321 SerAspSerSerAsnSerAsnCysIleThrTrpGluGlyThrAsnGlyGluPheLysMet 340
Db 1023 TCGGACAGCTCCAACTCCAACTGCATCCCTGCGGAGGCGCACAAATGGGAGTTCAAGATG 1082
QY 341 ThrAspProAspGluValAlaArgArgTrpGlyGluGlyLysSerLysProAsnMetAsn 360
Db 1083 ACAGACCTGATGAAGTGTCTCGGCGTTGGGAGAGAGAGAAAGCAACCTTAATGATGAC 1142
QY 361 TyrAspLysLeuSerArgAlaLeuArgTyrTyrTyrAspLysAsnIleMetThrLysVal 380
Db 1143 TATGACAAACTCAGCCGCGTCTGCTAGTACTATGACAAATAATATATGACTAAAGTT 1202
QY 380 ----- 380
Db 1203 CATGTAAGCGTATGCCATACAAATTTGATTTCCACGGAATGCTCAGGCCCTCAGGCT 1262

QY	381	HISPPROGLISERSEMETLYSTYRPROSERASPLEPROTYMETSERSEYR	400
Db	1263	CACCTCCAGAAATCATTCATGTACAAATACCATGACACTCCCTACATGAGTTCCTAC	1322
QY	401	HISGLYSARGTYRALATYRLYSPEASPRHEHISGLYLLEALAGLALALEUGLNPRO	420
Db	1322	-----	1322
QY	421	HISALAHISPROGLINLYSMEFASPHVEALAPROHISPROPROALALEUPROVALTHR	440
Db	1333	CATGACACACCCCAAGATGAATCTTGAGTCCCAATCCCTGCTTGGCCCTTAACC	1382
QY	441	SERSESRPHEPHALALAPROASPNROYTYRTPASNSERPROTHGLYGLYLLEYR	460
Db	1383	TCATCCAGCTTTTTTGGTCCGCCCTTAATCATCTGGAATTCACCAACTGAGACATCTAC	1442
QY	461	PROASNTPIARGLEUPROALAHISMETPROSERHISLEUGLYTHRTRYR	478
Db	1443	CCCAATACCAAGCTCCAGCTGCTCATATGCTTCCCACTTGGCACCTACTAC	1496
RESULT 3			
LOCUS	AB073080	2209 bp	linear
DEFINITION	Mus musculus Erg mRNA, mouse homolog of Human ets-related gene ERG,		
ACCESSION	AB073080		
VERSION	AB073080.1		
KEYWORDS	GI:16197544		
SOURCE	Mus musculus cDNA to mRNA, clone: Erg-3.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
	1		
	Ozawa, R., Noguchi, H., Taylor, T.D., Takeda, T., Hattori, M. and		
	Sakaki, Y.		
TITLE	Mus musculus Erg mRNA		
JOURNAL	Published Only in Database (2001)		
REFERENCE	2 (bases 1 to 2209)		
AUTHORS	Ozawa, R., Noguchi, H., Taylor, T.D., Takeda, T., Hattori, M. and		
	Sakaki, Y.		
TITLE	Direct Submission		
JOURNAL	Submitted (16-OCT-2001) Masahira Hattori, The Institute of Physical		
	and Chemical Research (RIKEN), Genomic Sciences Center (GSC);		
	1-7-22 Suehiro-chou, Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan		
	(E-mail: hattori@sc.riken.go.jp, url: http://hqp.gsc.riken.go.jp/,		
	Tel: 81-45-503-9111, Fax: 81-45-503-9170)		
FEATURES			
Source	Location/Qualifiers		
	1..2209		
	/organism="Mus musculus"		
	/db_xref="taxon:10090"		
	/chromosome="16"		
	/clone="Erg-3"		
	1..2209		
	/gene="Erg"		
	148..1608		
	/gene="Erg"		
	/note="homolog of human ets-related gene (ERG), transcript		
	variant 3"		
	/codon_start=1		
	/protein_id="BAB69950.1"		
	/db_xref="GI:16197545"		
	/translation="MIQTVDPDAHIREALSVESDQLFECAAYGTPLHAKTENTASS		
	SSDQGTSSKSPRYPOODMLSOAPARTIKMECPQVNGSRNPDECSYKRGKMGV		
	SPDVGSGSYMEKEHVPENMTNERRVIVPADPTLMSIDHYROWLEMAVKRGILL		
	DVYLIFONIDGKEIKCTKTDKDDPRLTPSYAAILLSHAIHRETRPLHLTSDVDKDR		
	LQNSPRLMARNTGCAATFIPNPISVYBATORITTRDPLETPRKSAMTSHSLTPD		
	SKAQPSPAVPKTEDQRPQDLPYQILGPTSSRLANGSGQIQLOMPLLELSDSSMS		
	NIETWEGTEGFKMTDEVAKRWGGEKSKRNMYDKLSRALRYTYKNIIMTKYHGRK		
	YAYFEDHGILAOALPHEPESLILKYPSDLPEMGSYNAHPQKNMFVSPHPALPYTSS		
	SFEASPPYNNMSPGTGIVPNTRLPSAHMPSHLGTYR"		
BASE COUNT	606 a 608 c 545 g 450 t		
ORIGIN			

Alignment Scores:			
Pred. No.:	2,34e-148	Length:	2209
Score:	2293.50	Matches:	431
Percent Similarity:	88.76%	Mismatches:	11
Best Local Similarity:	86.55%	Indels:	15
Query Match:	88.62%	Gaps:	41
DB:	10		3
US-09-902-772-4 (1-478) x AB073080 (1-2209)			
QY	2	AlaserThrIleLysGluAlaLeuSerValValSerGluAspGlnSerLeuPheGluCys	21
Db	172	GCACCTCATATTATGAGAGGCGCTTGCTGAGTGTGAGCGAGGACCAAGTCACATTATGAGTGT	231
QY	22	AlaIArgIleSerProHisIleuAlaIysThrGluMetThrIAsSerSerSerGluIArg	41
Db	232	GCCTACGGAAAGCCGACACCTGGGCTTAAACAGAAATGACCCCATCTCTTCCAGTACTAT	291
QY	42	GlyGlnThrSerIleLysMetSerProArgValProGlnGlnAspIleProLeuSerGlnProPro	61
Db	292	GGCCAGACATCCAGATGAGTCCAGAGTCCCTCAGCAGAGACAGTGGCTGTCTCAAGGCCCA	351
QY	62	AlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsnSer	81
Db	352	GCCAGGGTCACCAATCAAGATGGATGGACACCTCAAGTCAAGTGAATGTTCCAGGAACCTCA	411
QY	82	ProAspAspCysSerValAlaIAsGluGlyIleMetValSerSerSerSerAspAsnValGly	101
Db	412	CTGATGTGATGCTGAGTGGAAACAAAGGTGGGAACATGGTGGCCAGCCGGATCTGGGG	471
QY	102	MetaAsnIArgIleSerIArgMetGluGluValHisIleProProProAsnMetThrThrAsn	121
Db	472	ATGAGCTACAGGCGACACTACATGAGAGGAACATGCTCCGCTCCCAATATGACCAACAAAT	531
QY	122	GluArgArgValIleValProAlaAspProThrLeuTrpSerThrAspHisValArgGln	141
Db	532	GAGCGCAGAGTACGCTCCCTCGAGATCTTACCTGTGGAGCACAGACCATGCTCCGACAG	591
QY	142	TrpLeuGluTrpIleAlaValLysGluIuArgIleuProAspValAspIleIleuLeuPheGln	161
Db	592	TGGCTGGAGTGGGGGGGAAAGATATGAGCTCTCGATGTGAGCTCTTACTATTTTCAG	651
QY	162	AsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArgLeuThrPro	181
Db	652	AATATCATGGGAAGGAGCGTGTGAAACATACAAAGATATACCTTCCAGCGGCTCACAGCGG	711
QY	182	SerTrpAsnAlaAspIleLeuLeuSerHisLeuHisTrpLeuArgIleuThrProLeuPro	201
Db	712	AGCTACATGCGGCATTTCTTCTCTCAATCTCACTACTCTCAGAGAGATCCCTCCCTCCA	771
QY	202	HisLeuThrSerAspAspValAspLysAlaLeuGlnAsnSerProArgLeuMetHisAla	221
Db	772	CATGTGACTTCCATGATGACGTGTGATAGGCTTTCAAAAACCTCTCACGGTTAATGCATGCC	831
QY	222	ArgAsnThrGlyLysAlaThrPheIlePheProAsnThrSerValTrpProGluAlaThr	241
Db	832	AGAACACAGGGGGGTGACGCTTTATTTTCCCAATATCTCACTATATATCCGAGCTAGC	891
QY	242	GlnArgIleThrThrArgProAspLeuProTyrGluGlnAlaArgArgSerAlaTrpThr	261
Db	892	CAAAAGATTCAACTAGGCCAGATTTACCTTAAGACCTCCCAAGGAGTACACCTGGAGCC	951
QY	262	SerHisSerHisProThr---GlnSerLysAlaThrGlnProSerSerSerThrValPro	280
Db	952	GGCCACAGCCACCTCACCCCTCAGTCCAAAGCTGCTCAGGCATCTCTCTGACAGTGGCC	1011
QY	281	LysThrGluAspGlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThrIleSer	300
Db	1012	AAAACGTGAAGCCAGCGTCTCTCAGTTAAGATCTTTACACAGATCTCTGGAGCCAGCAAGTAC	1071
QY	301	ArgLeuAlaAsnProGlySerGlyGlnIleGlnLeuTrpGlnPheLeuLeuGluLeuLeu	320

Db	1072	GCSCITGCTAATCCAGAGTAGTGGCCAGATCCAGCTGTGGCAGTTCTCTGTCGAAC	CTCTG	1131
QY	321	SerAspSerSerAsnSerAsnCysIleThrTrpGluGlyThrAsnGlyGlnPheLysMet		340
Db	1132	TCAGACACTCCACATCCACACTGATCCACTGTGGAAAGCACACACGGGGAGTTCAAGATG		1191
QY	341	ThrAspProAspGluValAlaIarGargTrpGlyGluArgLysSerLysProAsnMetAsn		360
Db	1192	ACAGACCCGGAGAGAGTGGCTGGGGCTGGGGGAGAGAGAGAGACCCCAACATGTAGAC		1251
QY	361	TyrAspLysLysSerArgAlaLeuArgTyrTyrTyrAspLysAsnIleMetThrLysVal		380
Db	1252	TATGACAAAGCTACAGCCGGCCCTCCGGCTACTACTGACACAAACATCATGACCAAGGTG		1311
QY	381	HisProProGluSerSerMetLysTyrProSerAspLeuProTyrMetSerSerTyr		400
Db	1311	-----		1311
QY	401	HisGlyLysArgGlyTyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnPro		420
Db	1312	CACGGGAAGCGCTACGCCCTACAAAGTTGACTTCCACAGGAGATTGGCCAGGCCCTCAGGCC		1371
QY	420	-----		420
Db	1372	CACCTCTCTGAGTCTCCCTGTACAGATGACCCCTCCGACCTGCCATACATGGGCTCTAT		1431
QY	421	HisAlaHisProGlnLysMetAsnPheValAlaProHisProProAlaLeuProValThr		440
Db	1432	CACGCCACACCCCAAGAAATGAACTTTGTGTCTCCACACCTCCGGCTCTCCCAATCACA		1491
QY	441	SerSerSerPhePheAlaAlaProAsnProTyrTrpAsnSerProThrGlyGlyIleTyr		460
Db	1492	TCTTCAGATTCTCTTGTCTTCCTCCGAAACCAATGACTGGAATTCACGACGTGGGGCAATCTAC		1551
QY	461	ProAsnThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr		478
Db	1552	CCGAACACTAGGCTCCACGACGACCATATGCTCTTCACCTGGGACCTACTAC		1605
RESULT 4				
AY065661				
LOCUS	AY065661	1413 bp	mRNA	linear
DEFINITION	Gallus gallus erg isoform C-1-1 mRNA, complete cds; alternatively			
ACCESSION	AY065661			
VERSION	AY065661.1	GI:17887440		
KEYWORDS				
SOURCE	Gallus gallus.			
ORGANISM	Gallus gallus			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Archosauria; Aves; Neognathae; Galliformes; Phasianidae;			
TITLE	Phasianidae; Gallus.			
JOURNAL	1 (bases 1 to 1413)			
MEDLINE	Iwamoto,M., Higuchi,Y., Koyama,E., Enomoto-Iwamoto,M., Kurisu,K.,			
PUBMED	Yeh,H., Abrams,W.R., Rosenbloom,J. and Pacifici,M.			
AUTHORS	Transcription factor ERG variants and functional diversification of			
REFERENCE	chondrocytes during limb long bone development			
AUTHORS	J. Cell Biol. 150 (1), 27-40 (2000)			
TITLE	20531415			
JOURNAL	20531415			
MEDLINE	10893254			
PUBMED	10893254			
AUTHORS	2 (bases 1 to 1413)			
REFERENCE	Iwamoto,M., Higuchi,Y., Enomoto-Iwamoto,M., Kurisu,K., Koyama,E.,			
AUTHORS	Yeh,H., Rosenbloom,J. and Pacifici,M.			
TITLE	The role of ERG (ets related gene) in cartilage development			
REFERENCE	Osteoarthritis. Cartil. 9 Suppl A, S41-S47 (2001)			
TITLE	21535378			
JOURNAL	21535378			
MEDLINE	11680687			
PUBMED	11680687			
AUTHORS	3 (bases 1 to 1413)			
REFERENCE	Iwamoto,M., Higuchi,Y., Yeh,H. and Pacifici,M.			
AUTHORS	Direct Submission			
TITLE	Submitted (05-DEC-2001) Oral Anatomy and Developmental Biology,			
JOURNAL	Osaka University Faculty of Dentistry, 1-8 Yamadaoka, Suita, Osaka			
FEATURES	565-0871, Japan			
Location/Qualifiers				

source	1. 1413	/organism="Gallus gallus"	/db_xref="taxon:9031"	37. 1392	/note="alternatively spliced; transcription factor="	/codon_start=1	/product="erg isoform C-1-1"	/protein_id="FAL00889.1"	/db_xref="GI:17887441"	/translation="MASTIKELALSVSEEDSLFECAVSGPHLAKTEMTATSSSEYQGT SKMSPPQDDMLSPAPRYATIKMECPNPNVNSRNSPDDCSVAKGKMSSEDDNGM NCSYMEKHPPPPNMTNERRVLYVPADPLMTSDHYRQMLEAVKRYGLPVDILF QNDGKELCKMTKDFQRLTPSYNADLLSHLHLRERGAFTLEPNTSVYPEATQRT TRPDLPEQOARSRAMTSHSPQSKATQPESSVYKPEDRQPLDPTQILGPISSRIA NPSGQIQLOWFLELSDSSNSMCITWEGTNGFMTDEDEYVARRGERSKPNMY DKLSRALRYYYDKNIMTKVIGKRFAVEFDHGLAOLAPHPSSMWKYSDI.PYMS YHAIPQKMYNVAIPRALPYTPVTSSSFFAAPNPYNSPTGIGIYPNTRILPAHMPHSILGY
CDS										
	BASE COUNT	424 a	370 c	311 g	308 t					
	ORIGIN									
	Alignment Scores:									
	Pred. No.:	1.82e-146				1413				
	Percent Similarity:	2262.50				Matches:	430			
	Best Local Similarity:	86.35%				Conservative:	0			
	Query Match:	87.42%				Mismatches:	1			
	DB:	5				Indels:	67			
						Gaps:	3			
US-09-902-772-4 (1-478) x AY065661 (1-1413)										
OY	1	Meta1SerThrIleLeuGluAlaLeuSerValValSerGluAspGlnSerLeuPheGlu	20							
DB	37	ATGGCAACACATTTAAGGAAGCATTTACATGCGTGAAGACCACTCTTGTGGAG	96							
OY	21	CysAlaTYrglySerProHisIleuAlaLysThrGluMetThrAlaSerSerSerGlu	40							
DB	97	TGTGGCTAGCGATGCCCCACCTTCACAAGACAGAAATGACAGCCTCTCCACAGTAA	156							
OY	41	TYrglyGlnThrSerLysMetSerProArgValProGlnGlnAspTrpLeuSerGlnPro	60							
DB	157	TATGGGCAAACTCAAAATGAGCCCGCGCTCCCGCAGCGAGCTGGTTATCCACAGCC	216							
OY	61	ProAlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsn	80							
DB	217	CCGGCCAAAGTTACCATTAAGTAGGAGTAAACCAACAGTTAATGGGTCAAGAAAT	276							
OY	81	SerProAspAspCysSerValAlaLysGlyLysMetValSerSerSerAspAsnVal	100							
DB	277	TTCACCTGATGACTGACGGGTGGCAAAAGAGGAAATAGCTTACAGCTTCAGCAATGTT	336							
OY	101	GlyMetAsnTYrglySerTYrMetGlnGluLysHisIleProProProAsnMetThrThr	120							
DB	337	GGGATGAACTATGGAAGCTACATGGAAGAGAAGCAATATTCGCCCTCCAAATATGACAAC	396							
OY	121	AsnGluArgValIleValProAlaAspProThrLeuTrpSerThrAspHisValArg	140							
DB	397	AATGACAAAGAGTATGTGCGCACAATCTCTCAGTTATGAGACACAGACCAAGTACGG	456							
OY	141	GlnTrpLeuGlnTrpAlaValLysGlnTYrglyLeuProAspValAspIleLeuLeuPhe	160							
DB	457	CAGTGGCTGAGACTGGGCGACGTGAGAGAGTAAGTCTTCACAGACGTGGACATCTTTGTTC	516							
OY	161	GlnAsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArgLeuThr	180							
DB	517	CAGAACATTTGATGGGAAAGAGTGTGGTAAATGACCAAAAGATGACTCCAGAGACACG	576							
OY	181	ProSerTYrAsnAlaAspIleLeuLeuSerHisIleuHisTYrLeuArgGluTrpProLeu	200							
DB	577	CCGAGCTTAATACGACGAGATATCTCTCTGTCACACCTTACACTTCACAGAGAGAGA	630							
OY	201	ProHisLeuThrSerAspAspValAspLysValLeuGlnAsnSerProArgLeuMetHis	220							

Db 630 ----- 630
 QY 221 AAlaGAsnThrGlyAlaThrPheIleProAsnThrSerValTyrProGluAla 240
 Db 631 -----GGAGCCACTTTTATTTTCCAAATATCATCATGATTACCCAGAAAGCA 675
 QY 241 ThrGlnArgIleThrThrArgProAsnLeuProTyrGluGlnAlaArgArgSerAlaTrp 260
 Db 676 AGCAAGAGATTAACAACAAGGCCAGATTATCATCTTATGACACGAGAGATCAGCGTG 735
 QY 261 ThrSerHisSerHisProThrGlnSerLysAlaThrGlnProSerSerThrValPro 280
 Db 736 AGAGAGCAGACGATCCATCCATCGTCAAAAGCTACCAACCATCATCTTCAACAGGCC 795
 QY 281 LysThrGlnAspGlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSerSer 300
 Db 796 AAAACAGAAAGACCGCTCTCAGTTAGATCCCTATCATATTCCTTGAGCCAGCACACG 855
 QY 301 ArgLeuAlaAsnProGlySerGlyGlnIleGlnLeuThrGlnPheLeuGlnLeuLeu 320
 Db 856 CGCTTGGCAAAATCCAGAGGAGTGGGAGATACAGCTATGCGAGTTCCTACTGGAGCTTCG 915
 QY 321 SerAspSerSerAsnSerAsnGlyThrThrGlnGlyThrAsnGlyGlnPheLysMet 340
 Db 916 TCGGACAGCTCCCACTCCCACTGCATCACCCTGGAGAGGACAAATGGGAGTTCAAGATG 975
 QY 341 ThrAspProAspGluValAlaArgArgTrpGlyGluArgLysSerLysAspProAsnMetAsn 360
 Db 976 ACAGACCCCTGATGATGAGTGGCTGGCGCTGGGAGAGAGAAAAAGCAACCTATCATGATGAC 1035
 QY 361 TyrAspLysLeuSerArgAlaLeuArgTyrTyrTyrAspLysAsnIleMetThrLysVal 380
 Db 1036 TATGACAAACTGACCGCTGCTGCTGCTACTACTATGACAAAAATATATGACTAAAGTT 1095
 QY 380 ----- 380
 Db 1096 CATGGTAAAGCGTATGCTTCAAAATTTGATTTCCACAGGAATGGCTCAGGCCCTCAGCCT 1155
 QY 381 HisProProGlnSerSerMetLysTyrProSerAspLeuProTyrMetSerSerTyr 400
 Db 1156 CACCCCTCCGAAATCATCATGATACAAATACCCATCAGACCTCCCTACATGAGTTCTTAC 1215
 QY 401 HisGlyLysArgTyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnPro 420
 Db 1215 ----- 1215
 QY 421 HisAlaHisProGlnLysMetAsnPheValAlaProHisProProAlaLeuProValThr 440
 Db 1216 CATGCACACCCCAAGAGATGAACCTTGTAGCTCCCATCCCTGCTTGGCCCGTAAAC 1275
 QY 441 SerSerSerPhePheAlaAlaProAsnProTyrTyrPAsnSerProThrGlyLysIleTyr 460
 Db 1276 TCATCCAGCTTTTGTCTCCCTCATATCCATCTGGAATTCACCAACTGGAGCATCTTAC 1335
 QY 461 ProAsnThrArgLeuProAlaHisMetProSerHisLeuGlyThrTyrTyr 478
 Db 1336 CCCAATATACAGGCTGCCAGCTGCTCATATGCTTCCCATCTTGGACCTACTAC 1389
 RESULT 5
 E31253
 LOCUS E31253 1447 bp DNA linear PAT 18-JUN-2001
 DEFINITION Protein having cell calcifying inhibitory activity and gene
 E31253
 ACCESSION E31253
 VERSION E31253.1 GI:13025685
 KEYWORDS JP 1999075871-A/1.
 SOURCE unidentified.
 ORGANISM unidentified.
 REFERENCE 1 (bases 1 to 1447)
 AUTHORS Hiroyaasu,T., Yoshinobu,H., Marjio,P., Joel,R. and Helena,E.
 TITLE Protein having cell calcifying inhibitory activity and gene
 encoding the same

JOURNAL Patent: JP 1999075871-A 1 23-MAR-1999.
 CHUGAI PHARMACEUT CO LTD, HE TRUSTEES OF THE UNIVERSITY OF
 PENNSYLVANIA
 COMMENT OS Unidentified
 PN JP 1999075871-A/1
 PD 23-MAR-1999
 PE 26-MAY-1998 JP 1998166076
 PR 18-JUN-1997 US 08/678177,20-JUN-1997 US 60/050297 PI
 HIROYASU IWAMOTO, YOSHINOBU HIGUCHI, MARIJO PASHIYIKI, PI JOEL
 ROZENBUROOD,
 PI HELENA E
 PC C12N15/09,A61K48/00,C07K14/465,C07K16/18,C12Q1/68//A61K38/00,
 PC C12P21/02,
 PC C12P21/08,C12N15/00,A61K37/02
 CC Strandedness: Single;
 CC Topology: Linear;
 FH Key
 FT source 1..1447
 FT location/Qualifiers
 FEATURES location/Qualifiers
 source 1..1447
 /organism="unidentified"
 /db_xref="taxon:32644"
 BASE COUNT 440 a 374 c 317 g 316 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 1,88e-146 Length: 1447
 Score: 2262.50 Matches: 430
 Percent Similarity: 86.35% Conservative: 0
 Best Local Similarity: 86.35% Mismatches: 1
 Query Match: 87.42% Indels: 67
 DB: Gaps: 3
 US-09-902-772-4 (1-478) x E31253 (1-1447)
 QY 1 MetAlaSerThrIleLysGlnAlaLeuSerValValSerGluAspGlnSerLeuPheGlu 20
 Db 63 ATGGCAAGCAGCATATTAAGAGAGCATTTATCAGTGGTAAGTAAGACAGTCTTGTTGAG 122
 QY 21 CysAlaTyrGlySerProHisLeuAlaLysThrGlnMetThrAlaSerSerSerGlu 40
 Db 123 TGTGGCTTACGAGTCCGCCCACTTGCAGAAAGACAGAAATGACACCCCTTCCAGTGA 182
 QY 41 TyrGlyGlnThrSerLysMetSerProArgValProGlnGlnAspTrpLeuSerGlnPro 60
 Db 183 TATGGGCAAAACATCAAAAGATGAGCCGCGCTTCCCAAGAGACTGTTATCACAGCCC 242
 QY 61 ProAlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsn 80
 Db 243 CCGGCCAGAGTTTACATTAAGATGAGTGTAAACCCCAACAGGTTAATGGTCAAGGAAT 302
 QY 81 SerProAspAspCysSerValAlaLysGlyLysMetValSerSerSerAspAsnVal 100
 Db 303 TCACCTGATGACTGCGAGCGTGGCAAAAGAGGAAATGTTTACAGCTTACAGCAATGTT 362
 QY 101 GlyMetAsnTyrGlySerTyrMetGluGluLysHisIleProProProAsnMetThrThr 120
 Db 363 GGGATGAAGTATGAGAACTACATGAGAAAGAGACATATTCGCGTCCCAATATGACAAAC 422
 QY 121 AsnGluArgArgValIleValProAlaAspProThrLeuTrpSerThrAspHisValArg 140
 Db 423 AATGAACGAAGATTATTTGTCCAGAGATCTTACTTATGAGGACACATCATGACG 482
 QY 141 GlnTrpLeuGlnTrpAlaValLysGluTyrGlyLeuProAspValAspIleLeuPhe 160
 Db 483 CAGTGGCTGGAAGTGGCAATGAAGAGATGCTTCCAGAGCTGACATCTTGTGTTC 542
 QY 161 GlnAsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArgLeuThr 180
 Db 543 CAGAACATTGATGGGAAGAGTGTGTAAATAGACCAAAAGATGACTCCAGAGACTCAG 602
 QY 181 ProSerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGlnThrProLeu 200

```

|||||
Db 603 CCGAGCTAATACGACAGTATTCCTCTCTCACCCTACACTACTCAGAGAGAGA----- 656
Qy 201 ProHisLeuThrSerAspValAspValAlaLeuGlnAsnSerProArgLeuMetHis 220
Db 656 ----- 656
Qy 221 AlaArgAsnThrGlyGlyAlaThrPheLeuPheProAsnThrSerValTyrProGluAla 240
Db 657 -----GGAGCAGCTTTTATTTTCCAAATATCAGTTCACCGAGGAGCA 701
Qy 241 ThrGlnArgThrThrArgProAspLeuProTyrGlnGlnAlaArgAsnSerAlaTyr 260
Db 702 ACCGAAAGAAATACACAAAGCCAGATTACCTTATGAGCAAGGAGAGATCAGCGTGG 761
Qy 261 ThrSerHisSerHisProThrGlnSerGlyAlaThrGlnProSerSerThrValPro 280
Db 762 ACAGATCAGCCAGCCATCCACTCAGTCAAAAGCTAACCAACCATCATCTTCACAGTCCC 821
Qy 281 LysThrGlnAspGlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSerSer 300
Db 822 AAAACAGAAACACAGCGCTCCTCAGTTGATCTTATCAGATCTTGGAGCCAGCAGCAGC 881
Qy 301 ArgLeuAlaAsnProGlnSerGlyGlnIleGlnLeuThrPheLeuLeuGlnLeuLeu 320
Db 882 CGTCTGCAATCCAGGAGAGTGGCAGATACAGCTATGAGCACTTCACGAGAGCTTCG 941
Qy 321 SerAspSerSerAsnSerAsnCysIleThrTrpGluGlyThrAsnGlyGluPheLysMet 340
Db 942 TCGGACAGCGTCCAACTCCAACTGCATCAGCTGGAGAGGCAAAATGGGAGATTCAGATG 1001
Qy 341 ThrAspProAspGluValAlaArgArgTrpGlyGluArgLysSerLysProAsnMetAsn 360
Db 1002 ACAGACCTCATATGAAGGCGCTCGGCGTGGGAGAGAGAAAGCAAACTTAACATGAC 1061
Qy 361 TyrAspLysLeuSerArgAlaLeuArgTyrTyrTyrAspLysAsnIleMetThrLysVal 380
Db 1062 TATGACAAACTCAGCCGCTGCACTTGCCTACTACTATGACAAATAATATATGACTAAAGTT 1121
Qy 380 ----- 380
Db 1122 CATGTAAACGCTATGCTACAAATTTGATTTCCACGAGATTCGCTCAGGCCCTCCAGCCT 1181
Qy 381 HisProGlnLeuSerSerMetTyrLysTyrProSerAspLeuProTyrMetSerSerTyr 400
Db 1182 CACCTCCCAAAATCATCCATGATCAATATACCATCAGACCTCCCTCAGATGAGTTCTTAC 1241
Qy 401 HisGlyLysArgTyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnPro 420
Db 1241 ----- 1241
Qy 421 HisAlaHisProGlnLysMetAsnPheValAlaProHisProProAlaLeuProValThr 440
Db 1242 CATGACACACCCCAAGAGATGACTTTGTAGCTCCCATGCCCTGCTTGGCCCGTACAC 1301
Qy 441 SerSerSerPhePheAlaAlaProAsnProTyrTyrPheAsnSerProThrGlyLysIleTyr 460
Db 1302 TCATTCAGAGCTTTTGTGCTGCCCTATATCCATCTGAGATTCACCAACTGGAGGATCTAC 1361
Qy 461 ProAsnThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr 478
Db 1362 CCCAAATACAGGCTGCCAGCTGCTCATATGCTTCCATCTTGGCAGCTACTTAC 1415

RESULT 6
AB073079 2133 bp mRNA linear ROD 17-OCT-2001
LOCUS AB073079 Mus musculus Ery mRNA, mouse homolog of human ets-related gene ERG,
DEFINITION transcript variant 2, complete cds.
ACCESSION AB073079.1 GI:16197542
VERSION AB073079.1
KEYWORDS Mus musculus cDNA to mRNA, clone: Ery-2.
SOURCE Mus musculus
ORGANISM Mus musculus

```

```

REFERENCE
AUTHORS
1 Ozawa, R., Noguchi, H., Taylor, T. D., Takeda, T., Hattori, M. and Sakaki, Y.
TITLE
JOURNAL
2 Published only in Database (2001)
REFERENCE
AUTHORS
2 (bases 1 to 2133)
Ozawa, R., Noguchi, H., Taylor, T. D., Takeda, T., Hattori, M. and Sakaki, Y.
TITLE
JOURNAL
Submitted (16-OCT-2001) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC), 1-7-22 Suenho-cho, Tsukuba, Ibaraki, Japan (E-mail: hattori@gsc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/, Tel: 81-45-503-9111, Fax: 81-45-503-9170)
FEATURES
source
1. 2133
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="16"
/clone="Erg-2"
1. 2133
/gene="Erg"
149. 1340
/note="homolog of human ets-related gene (ERG), transcript variant 2"
/codon_start=1
/db_xref="GI:16197543"
/translating="MOTVDPDPAHKEALVSSEDOSEFECAGTPHILAKTEMTASS
SSDYQTSKMSPPVPODMLSQAPAVTTIMECPNSQVNSRSPDEC SYNKGKMG
SPDTVMSGYSYEEKHVPPTNERIVYADPTIMSTDRVOMLEAVKVEYGL
DVTDLFONIDKEIKMRDDDFORLPSYNAIILSHLYLETPHLSDDVYKA
LONSPRLMARNTGGAFTFPNSVPEAQRTTTPRAOPSSAPKTKDOROLDP
YOTLPTSSHLANSGGQIOWOLFLELSDSNKCTWEGTNGEKMDPDEVAR
WGRKRSKPMNVDKLSRALRYDDKIMTKVHKRAYKDFGILQALOPHPEBSL
YKPSDLPYMSYHAHPDKMNEVSPHPALPYSSSFASPNPNWNSPTGIVPNRL
PASHMPSHLGTYR"
BASE COUNT 587 a 579 c 531 g 436 t
ORIGIN
Alignment Scores:
Pred. No.: 5.35e-141 Length: 2133
Score: 2186.00 Matches: 413
Percent Similarity: 85.31% Conservative: 11
Best Local Similarity: 83.10% Mismatches: 11
Query Match: 84.47% Indels: 62
DB: 10 Gaps: 3
US-09-902-772-4 (1-478) x AB073079 (1-2133)
Qy 2 AlaSerThrIleLysGluAlaLeuSerValValSerGlnAspGlnSerLeuPheGluCys 21
Db 173 GCACTTCATATTTAAGGAGCGCTTGTGAGTGAGCAGGACCATGATATTTGAGAGT 232
Qy 22 AlaTyrGlySerProHisLeuAlaLysThrGlnMetThrAlaSerSerSerGlyTyr 41
Db 223 GCCACGGAACGCCACACCTGGCTTAACAGACAGATCCGATCTTTCACAGTACTAT 292
Qy 42 GlyGlnThrSerLysMetSerProArgValProGlnGlnAspTyrLeuSerGlnProPro 61
Db 293 GGCACAGACATCAAGATGATGCCAGAGATCCCTCAGAGAGCTGGTGTCCAGGCCCA 352
Qy 62 AlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsnSer 81
Db 353 GCCAGGCTCACCAAGAGAGAGTGCACACCTAGAGTGAATGGTCCAGCACTCA 412
Qy 82 ProAspAspCysSerValAlaLysGlyGlyLysMetValSerSerSerAspAsnValGly 101
Db 413 CCGATAGAGTCACTGTGACCAAAAGGTGGAGAGATGTTGGGACAGCCCGATCTGTGGG 472

```

```

OY 102 MetAsnTyrGlySerTyrMetGluGlyHisIleProProProAsnMetThrTrpAsn 121
    |||
Db 473 AGGAGTACGGAGCTACATGAGGAGGAGGAGCATGCGCGCTCCCATATGACCAAAAT 532
OY 122 GluArgArgValIleValProAlaAspProThrLeuTrpSerThrAspHisValArgGln 141
    |||
Db 533 GACGGAGAGTGAATGCTCCCTGCAGATCTCTACTGTGGAGCACAACCATGTGCCACAG 552
OY 142 TrpLeuGluTrpAlaValLysGluTyrGlyLeuProAspValAspIleLeuLeuPheGln 161
    |||
Db 593 TGGCTGGAGTGGCGGCTGAAGAAATAGCGCTCCCTGATGGAGCATCTTACTATATTTCAG 652
OY 162 AsnIleAspGlyLysGluLeuGlySerMetThrLysAspAspPheGlnArgLeuThrPro 181
    |||
Db 653 AATATGATGAGGAGAGGAGCTGTGCAGAGATGACAAAGAGTACTCCAGCGGCTCACGGCG 712
OY 182 SerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGluThrProLeuPro 201
    |||
Db 713 AGCTACAAATGCCGACATCTTCTCTCACATCTCCACTACTCTGAGAGACTCCCTTCCA 772
OY 202 HisLeuThrSerAspAspValAspLysAlaLeuGlnAsnSerProArgLeuMetHisAla 221
    |||
Db 773 CATCTACTCTCGATACGCTTGATAGGCTTACAAACTCTCCACGGTTAATGATGCC 832
OY 222 ArgAsnThrGlyLysAlaThrPheIlePheProAsnThrSerValTyrProGluAlaThr 241
    |||
Db 833 AGAACAACAGGGGGGTCACCTTTATTTCCAAATATCTGATATATCCGAAAGCTAG 892
OY 242 GlnArgIleThrThrArgProAspLeuProTyrGluGlnAlaArgSerAlaTrpThr 261
    |||
Db 893 CAAAGAAATACAACTAGGCCA----- 913
OY 262 SerHisSerHisProThrGlnSerLysAlaThrGlnProSerSerSerThrValProLys 281
    |||
Db 914 -----GCTGCTCAGCCATCTCCCTGCTGAGGCCAAA 946
OY 282 ThrGluAspGlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSerSerArg 301
    |||
Db 947 ACTGAGACACAGGCTCTAGTATGATCTTACCAATCTGAGTATATCCGAAAGCTAG 1006
OY 302 LeuAlaAsnProGlySerGlyGlnIleGlnLeuTrpGlnPheLeuLeuGluLeuLeuSer 321
    |||
Db 1007 CTTCGTAATCCAGTAGTAGGCCAGATCCAGCTGTGGCACTTCCTGCTGCAACTCCGTGA 1066
OY 322 AspSerSerAsnSerAsnGlyIleThrTrpGluGlyThrAsnGlyGluPheLysMetThr 341
    |||
Db 1067 GACAGCTCCAACTCCAAAGCATCACTCTGGAGGAGCACCAGCGGAGTTCAAGATGACA 1126
OY 342 AspProAspGluValAlaIleArgArgTrpGlyGluArgLysSerLysProAsnMetAsnTyr 361
    |||
Db 1127 GACCCGGAGAGAGGTGGCTGGCGTGGGGGAGAGAGAGCAAGCCCAACATGAACTAT 1186
OY 362 AspLysLeuSerArgAlaLeuArgTyrTyrTyrAspLysAsnIleMetThrLysValHis 381
    |||
Db 1187 GACAAAGCTAGCGCGCGCTCCGCTACTACTACGACAAAAAATCATCTGACCAAGGCG--- 1243
OY 382 ProProGluSerSerMetLysTyrProSerAspLeuProTyrMetSerSerTyrHis 401
    |||
Db 1244 -----CAC 1246
OY 402 GlyLysArgTyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnPro--- 420
    |||
Db 1247 GGGAAAGCGCTACGCTTACAAAGTTTACTTCCACGGGATTTGCCAGGCGCTGACGCCAC 1306
OY 421 -----His 421
Db 1307 CCTCTGAGTCTCCCTGACAGATACCCCTCCGACCTGCCATACATGGGCTCTCTATCAC 1366
OY 422 AlaHisProGlnLysMetAsnPheValAlaProHisProProAlaLeuProValThrSer 441
    |||
Db 1367 GCCACACCCCAAGAAATGAACTTTGTGTCTCCACCCCTCCGCTCTCCACATCATATCT 1426
OY 442 SerSerPhePheAlaIleAlaProAsnProTyrTrpAsnSerProThrGlyLysIleTyrPro 461

```

```

Db 1427 TCCAGTTTCTTTCCTTCCCGGAGCCCATACTGGAATTCACCGGAGGAGCATCTACCG 1486
OY 462 AsnThrArgLeuProAlaHisMetProSerHisLeuGlyThrTyrTyr 478
    |||
Db 1487 AACACTAGGCTCCACGACCAACCATATGCTCTTCACTGCGGACCTACTTAC 1537
RESULT 7
HUMERG2
LOCUS HUMERG2 3166 bp ss-RNA linear Pri 08-NOV-1994
DEFINITION Human erg2 gene encoding erg2 protein, complete cds.
ACCESSION M17254
VERSION M17254.1 GI:182186
KEYWORDS erg 2 protein.
SOURCE Human cell line COLO 320, cDNA to mRNA, clone lambda 12.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
1 (bases 1 to 3166)
AUTHORS Rao,V.N., Paps,T.S. and Reddy,E.S.
TITLE erg, a human ets-related gene on chromosome 21: alternative
JOURNAL splicing, polyadenylation, and translation
MEDLINE Science 237 (4815), 635-639 (1987)
PUBMED 87263429
FEATURES
source Location/Qualifiers
1..3166
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="21q22.3"
257..1645
/gene="ERG"
257..1645
/gene="ERG"
/codon_start=1
/protein_id="AA52398.1"
/db_xref="GI:182187"
/db_xref="DB:G00-119-884"
/translation="MIVPDPDAHIKEALSVSHDSLFECAYGTPHIAKTEMTASS
SSDVGOTSKMSPRVPODMLSOPPARVTIKECNSOVNGSRNSPDECSVAKGKMVG
SPDVTGMVSGYMEERHMPDPMTNERRVTPADPTLMSRDHVRHLEMAVKYEGGLP
DYNILLPONTIGKEIKETKTKTRDPQRLTPTSYNADILSLHRLRETPHLVLSVDVKA
LONSRLMHAARTDLPTEPRKSAWTEGHPTPOSKAQSPSTVPKTEDRPLDQPY
QILGPTSSRLNLPKSGQIQIQWQLEILSDSNSCITMBETNGFEFKTDDDEVARRV
GERKSKPMNMDKLSRLARYYYDKNIMTKVIGKRAYAFDFHGLAQAQPPPESSLY
KVPSDLPYMGSVYHAPQKMFVAPHPALPYTSSSEFAAPPYMNSPIGILYPTRLP
TSHMPSHICITY"
repeat_region 291..298
/feature="8-bp repeat sequence"
repeat_region 509..516
/feature="8-bp repeat sequence"
polyA_signal 3143..3149
/feature="putative"
BASE COUNT 915 a 770 g 713 t
ORIGIN 227 bp upstream of PvuII site; chromosome 21.
Alignment Scores:
Pred. No.: 2,07e-140 Length: 3166
Score: 2180.50 Matches: 413
Percent Similarity: 84.74% Conservative: 9
Best Local Similarity: 82.93% Mismatches: 11
Query Match: 84.25% Indels: 65
DB: 9 Gaps: 4
US-09-902-772-4 (1-478) x HUMERG2 (1-3166)
OY 2 AlaSerThrIleLysGluAlaLeuSerValYalSerGluAspGlnSerLeuPheGluLys 21
    |||
Db 281 GCAGCTCATATCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 340
OY 22 AlaTyrGlySerProHisLeuAlaLysThrGluMetThrAlaSerSerSerSerLysTyr 41
    |||

```

Db 341 GCCACGGAACGCCACCTGGCTAAGACAGATGACCCGCTCTCTCCAGGACTAT 400
 Qy 42 GlyInThrSerLysMetSerProArgValProGlnInAspTyrPheSerGlnProPro 61
 Db 401 GGACAGACTTCGCAAGATGAGCCACGGCTCCCTCAGCAGATTTGGCTGTCCAAACCCCA 460
 Qy 62 AlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsnSer 81
 Db 461 GCCGGGTCTACCAATCAAAATGGAATGTACCTGTGCGCAGGATGGCTCCAGGAACTCT 520
 Qy 82 ProAspAspCysSerValAlaLysGlyGlyLysMetValSerSerSerAspAsnValGly 101
 Db 521 CCTATGATTAATGCACTGTGGCCAAAGCGGAGATGGTGGCACCACAGACACCGTGGG 580
 Qy 102 MetAsnTyrGlySerTyrMetGluGlnLysHisIleProProProAsnMetThrThrAsn 121
 Db 581 ATGAATACAGGCGACTACATGAGAGAGACATGCGACCCCAACATGACACAGCAGMAC 640
 Qy 122 GluArgArgValIleValProAlaAspProThrIleuTyrPheSerThrAspHisValArgGln 141
 Db 641 GAGCGAGAGTATATGCGCAGCAGATCTACCTATGAGATACAGACCATAGTGGCAG 700
 Qy 142 TyrPheGluTyrPheValValLysGluTyrGlyLysProAspValAspIleLeuLeuPheGln 161
 Db 701 TGGCTGAGTGGCGGCGGGAAGATATGGCTTCACAGCGTCAACATCTTGTATTCCAG 760
 Qy 162 AsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArgLeuThrPro 181
 Db 761 AACATCATGAGGAGAAAGAACTGTGCAGATGACCAAGGCGACTTCGAGAGCTCACCC 820
 Qy 182 SerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGluThrProLeuPro 201
 Db 821 AGCTACACAGCGCGCACTCTCTCTCATCTCCATCTCAGACAGAGACTCCCTCTCA 880
 Qy 202 HisLeuThrSerAspAspValAspLysValAlaLeuGlnAsnSerProAlaGlyMetHisAla 221
 Db 881 CATTTGACTTCAGATGATGTTGATTAAGCCCTTACAAACTCTCCACGGTTATATCATGCT 940
 Qy 222 ArgAsnThrGlyGlyAlaThrPheIlePheProAsnThrSerValTyrProGluAlaThr 241
 Db 941 AGAACAACA----- 949
 Qy 242 GlnArgIleThrThrArgProAspLeuProTyrGluGlnAlaArgSerAlaTyrThr 261
 Db 950 -----GATTTACCATATGAGGCCGCCCGCAGAGATCATCGCTTGAGCC 988
 Qy 262 SerHisSerHisProThr-----GlnSerLysAlaThrGlnProSerSerSerThrValPro 280
 Db 989 GGTACACGGCCACCCACGCCCACTGCAAAAGCTCTCAACCATCTCTCCACAGTCCC 1048
 Qy 281 LysThrGluAspGlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSerSer 300
 Db 1049 AAAATGAAACACCGCTCTCTAGTATGATTCCTATCAGATTCTTGGACCAACAAATGAC 1108
 Qy 301 ArgLeuAlaAsnProGlySerGlyGlnIleGlnLeuTyrGlnPheLeuLeuGlnLeuLeu 320
 Db 1109 CGCCTTCAATCCAGGACAGTGGCAGATCCAGCTTGGAGATCCCTCGGAGCTCTCG 1168
 Qy 321 SerAspSerSerAsnSerAsnCysIleThrTyrPheGluTyrHisAsnGlyLysPheLysMet 340
 Db 1169 TCGGACAGCTCCAACTCCAGCTGATCATCTCGGAAGGACCAACCGGAGATTCAAGATG 1228
 Qy 341 ThrAspProAspGluValAlaArgArgTyrGlyGluArgLysSerLysProAsnMetAsn 360
 Db 1229 ACGGATCCGACGAGGTGGCGCGCTGGGAGAGCGGAAGACCAACCAACATGATGAC 1288
 Qy 361 TyrAspLysLysSerArgAlaLeuArgTyrTyrTyrAspLysAsnIleMetThrLysVal 380
 Db 1289 TACATATAGCTCAGCCGCGCTCTGTTACTACTATGACAAAGACATCATGACCAAGGTC 1348
 Qy 381 HisProGluLysSerMetTyrLysTyrProSerAspLeuProTyrMetSerSerTyr 400
 Db 1348 ----- 1348

Qy 401 HisGlyLysArgTyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnPro 420
 Db 1349 CATGGGAAGCGCTACGCTTACAAAGTTGACTTCACAGGAGATGGCCAGCCCTCCAGGCC 1408
 Qy 420 ----- 420
 Db 1409 CACCCCGGAGTATCTCTGTACAGTACCCTCAGACCTCCGATACGCTCTCAT 1468
 Qy 421 HisAlaHisProGlnLysMetAsnPheValAlaProHisProProAlaLeuProValThr 440
 Db 1469 CACGCCCAACCCAGAAATATACTTGTGGCGCCCACTCTCCAGCCCTCCGCTGACA 1528
 Qy 441 SerSerSerPhePheAlaAlaProAsnProTyrTyrPheAsnSerProThrGlyLysIleTyr 460
 Db 1529 TCTTCAGATTTTTTGTGCGCCCAACCATAGTATGATTCACCAACTGGGGATATATAC 1588
 Qy 461 ProAsnThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr 478
 Db 1589 CCCAACACTAGGCTCCGCCACGACCATATGCTCTCATGTGGCATTTACTAC 1642
 RESULT 8
 LOCUS XLAJ4125 1798 bp mRNA linear VRT 21-JAN-2000
 DEFINITION Xenopus laevis erg gene (erg_A).
 ACCESSION AJ224125
 VERSION AJ224125.1 GI:5420045
 KEYWORDS ERG gene; transcription factor.
 SOURCE Xenopus laevis.
 ORGANISM Xenopus laevis.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae;
 Xenopodinae; Xenopus.
 1 (bases 1 to 1798)
 Balitzinger,M., Mager-Heckel,A.M. and Remy,P.
 TITLE Xl erg: expression pattern and overexpression during development
 JOURNAL Dev. Dyn. 216 (4-5), 420-433 (1999)
 MEDLINE 20099678
 PUBMED 10633861
 REFERENCE 2 (bases 1 to 1798)
 Balitzinger,M.
 TITLE Direct Submission
 AUTHORS Submitted (24-FEB-1998) Balitzinger M., UPR9005, MMDCD, Centre
 JOURNAL National de la Recherche Scientifique, 15, RUE Rene Descartes,
 67084, FRANCE
 FEATURES
 source location/Qualifiers
 1..1798
 /organism="Xenopus laevis"
 /db_xref="taxon:8355"
 /clone="lambda ZAPII-erg A"
 /clone_1b="lambda ZAPII"
 /dev_stage="stage 30 embryo"
 321..1778
 /gene="erg"
 321..1778
 /function="transcription factor"
 /codon_start=1
 /protein_id="CAB6566.1"
 /db_xref="GI:5420046"
 /db_xref="SPTREMBL:O9W700"
 /translation="MOHMAFIPYEDSNALSVSDDSLFECTYGTPIHTRKEMTA
 SSSDVGOTSKMSPRYQODMLSOEPSRYVTKMCNPSOYNGSRSSPDCSICKSKM
 GGSDDVGMVSGSYIEKHHSPTMPTNERRYVPAADPTLMSYDHRQMLEAKIEKG
 LRPVDVLIRONIDKECKRTKEDFHLRTSYADLLSHLHLREPLHLTSDVD
 KALQNSRLHAKNTGASFIIPNSSVYODANORIPRSQDLYEPRSRSAMTHNPAP
 SKASOPSTVPMKTEDPPLDPOILGPTSSRLANPSGQIOLOPFLBELTSDSN
 CITWEGNGEKFMTDDEVARWGERKSKPMNYDKLSRLRYVDKNIIMTVHGRKY
 AYKEDFHGIAOLQHPPESTMYKPESELPMYSYHNPQKMFVAHPHAPALVTSS
 FFAAPNAYMSPGSIYPMNRRLASHMSHLGITY"
 BASE COUNT 561 a 429 c 358 g 450 t
 ORIGIN

CDS 63..1430
/gene="Vespl14"
/note="similar to human erg1"
/product="vascular endothelial cell specific protein 14"
/protein_id="BAB62744.1"
/db_xref="GI:15128489"
/translation="MASTIKREALSVTSKQSLFECAVYGPPIHAKTMTASSSDYQOT
SKMSPVPOQDWLSOPPARVTKMCNPSQVNSRNSPDESVTKGKAVGSPDIVGM
SYGSMEEKHMPNMTNREVRIVPADPLTMSTNRQMLEVAEYGLADYDLF
ONIDKEKCMKEDFORLEPYSNADILLSHLLEYRPLPHLSDVDKALQNSRL
MHAHNDLPYEPFRSTWTPSHPTPOSKAOPSPEVTKTEORPOLDPYOLIGPTS
SRLANGSSQIQLOMOLLELSSSNSKNTMGTEGKERMDPDVARMRGKRKP
NMNIDKSLRALRTIDKNIMTKVHKRIRAKFEDHGIQAQLQHPPESSLYIKPSLP
YMGSYHTHPQKNFVAPHPALPVTSSSEFATPNPYMNSPTGIGLYNTRILPASHMESH
LGTYV"

BASE COUNT 510 a 522 c 426 g 350 t
ORIGIN

Alignment Scores:
Pred. No.: 1,68e-140 Length: 1808
Score: 2177.50 Matches: 412
Percent Similarity: 84.37% Conservative: 9
Best Local Similarity: 82.57% Mismatches: 13
Query Match: 84.14% Indels: 65
DB: 10 gaps: 4

US-09-902-772-4 (1-478) x AB031088 (1-1808)

QY 1 MetAlaSerThrIleuValAlaLeuSerValValSerGluAspGlnSerLeuPheGlu 20
DB 63 ATGGCCAGCACTATTAAAGGAGGCGCTTGTCAGTGTGAGCAAGCCAGTCACTATTTGAG 122
QY 21 CysAlaIleuSerProHisLeuAlaIleuThrGluMetThrAlaSerSerSerGlu 40
DB 123 TGTGCGTACGGAAGCGCACACCTGGCTTAACAGAAATGACCGGATCTCTCCAGTGAC 182
QY 41 TyrGlyGlnThrSerIleuMetSerProAlaValProGlnGlnAspTyrPheSerGlnPro 60
DB 183 TATGGCCAGCACTCAAGAGAGTCCAGATGCCACACAGCAGCTGCTGCACAGCC 242
QY 61 ProAlaIleuValThrIleuMetGluCysAsnProAsnGlnValAsnGlySerArgAsn 80
DB 243 CCAGCCAGGCTCACCATTCAAGATGAGTGCACACCTTAGCAGGTGACAGGTTCCAGGAGC 302
QY 81 SerProAspAspCysSerValAlaIleuGlyIleuMetValSerSerSerAspAsnVal 100
DB 303 TCACCTGACGATGCACTGTGACCAAGAGTGGGAGATGGTGGGACACCTGATACCTGTG 362
QY 101 GlyMetAsnTyrGlySerTyrMetGluGluIleuHisIleProProProAsnMetThrThr 120
DB 363 GGATGTGACGACGCGCCTCATGAGAGACACACATGCCGCCCAATATGACACAG 422
QY 121 AsnGluIleuArgValIleValProAlaAspProThrLeuTyrPheSerThrAspHisValArg 140
DB 423 AATGTAACGAGAGATGCTCCCTGCAATCTCACTGTGGAGCAACAGCAATGTCGG 482
QY 141 GlnTyrPheGlnTyrTrpAlaValIleuGlyIleuProAspValAspIleuLeuPhe 160
DB 483 CAGTGGCTGGAGTGGGAGGAAAGATACGGCTCTCAACGCGGAGCGCTTTATTATT 542
QY 161 GlnAsnIleuAspGlyIleuGluLeuCysIleuMetThrIleuAspAspPheGlnArgLeuThr 180
DB 543 CAGAAATGTGACGGGAGGAGCTGTCCAGATGACAAAGATGACTTCCAGAGGCTCACT 602
QY 181 ProSerIleuAsnIleuAspIleuLeuSerHisIleuHisTyrLeuArgGlnTyrProLeu 200
DB 603 CCGAGCTACAGATGCGCAGATCTTCTCTCACTCTCACTACCTCAAGAGAGACTCCCTT 662
QY 201 ProHisIleuThrSerAspAspValAspIleuAlaLeuGlnAsnSerProArgLeuMetHis 220
DB 663 CCACATCTGACATCCGAGAGAGTGTGATAGGCTTTACAAAACCTCTCCACGGTTAATGAT 722

QY 221 AlaArgAsnThrGlyGlyAlaThrPheIlePheProAsnThrSerValTyrProGlnAla 240
DB 723 GCTTGAAGAAACACA----- 734
QY 241 ThrGlnArgIleThrThrArgProAspLeuProTyrGlnGlnAlaArgArgSerAlaTyr 260
DB 735 -----GATTTACCTTATAGCCTCCACAGAGATCAACCTGG 770
QY 261 ThrSerHisSerHisProThr-----GlnSerLysAlaThrGlnProSerSerSerVal 279
DB 771 ACCGGCAGCAGCCACCCACCCCTCAGTCCAGCAAAAGCTGCCAGCATCTCCCTCAGAGTG 830
QY 280 ProLysThrGluAspGlnArgProGlnLeuAspProTyrGlnIleuGlyProThrSer 299
DB 831 CCCAAACCTGAAGACCGAGCGTCCCTAGTATGATCTTACAGATCCGAGATCCGGGACCCAGT 890
QY 300 SerArgLeuAlaAsnProGlySerGlyGlnIleGlnIleuTyrPheIleuLeuGlnLeu 319
DB 891 AGTGCCTTGCTAATCCAGAGTGTGGCCAGATCCAGCTGTGGCAGTCTGTGAACCTC 950
QY 320 LeuSerAspSerSerAsnSerAsnGlyIleThrTyrGluGlyThrAsnGlyIlePheLys 339
DB 951 CTGTCTACAGCTCCCACTCCCACTGCATCACCCTGGGAGGACACCAACGGGAGTTCAAG 1010
QY 340 MetThrAspProAspGluValAlaAlaArgArgTyrGlyGluArgLysSerLysProAsnMet 359
DB 1011 ATGACAGACCCGATGAGTGTGCTCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1070
QY 360 AsnTyrAspLysIleuSerArgAlaLeuArgTyrTyrTyrAspLysAsnIleuMetThrLys 379
DB 1071 AACTATACAAACTCAGACCGGCGCTCCGCTACACTACGACCAAAACATCATACCAAG 1130
QY 380 ValHisProProGlnIleuSerMetTyrLysTyrProSerAspLeuProTyrMetSerSer 399
DB 1131 GTG----- 1133
QY 400 TyrHisGlyLysArgTyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGln 419
DB 1134 ---CACGGGAAGCGCTATGCTTACCAAGTTGACTTCACAGGGAGTTGCCAGCCCTGCAG 1190
QY 420 Pro----- 420
DB 1191 CCCCATCCCCCGAGTGTCTCCCTTACAGATACCCCTCCGACCTGCCGTACATGGGCTCC 1250
QY 421 ---HisAlaHisProGlnLysMetAsnPheValAlaProHisProProAlaLeuProVal 439
DB 1251 TATACACACCCACCCCGAAGATGATGCTGTGCTCCACCCCTCGGGCCCTCCAGTGC 1310
QY 440 ThrSerSerSerPhePheAlaAlaProAsnProTyrTyrPheAsnSerProThrGlyIle 459
DB 1311 ACATCTTCCAGTTCTTGTGTACCCCGAACCCCATACATGGAATTCGCCGACATGGGGCATGC 1370
QY 460 TyrProAsnThrArgLeuProAlaAlaHisMetProSerHisLeuGlyTyrThrTyr 478
DB 1371 TACCGGAACACTAGGCTCCAGCCAGCATATGCTCTCACTGCGGACCTACTAC 1427

RESULT 10
AB073078 2131 bp mRNA linear ROD 17-OCT-2001
LOCUS
DEFINITION
Mus musculus Erg mRNA, mouse homolog of Human ets-related gene ERG,
transcript variant 1, complete cds.
AB073078
AB073078.1 GI:16191716
VERSION
KEYWORDS
SOURCE
ORGANISM
Mus musculus cDNA to mRNA, clone: Erg-1.
Mus musculus
Eukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathia; Muridae; Murinae; Mus.
1
REFERENCE
Ozawa, R., Noguchi, H., Taylor, T. D., Takeda, T., Hattori, M. and
Sakaki, Y.
TITLE
Mus musculus Erg mRNA
JOURNAL
Published Only in Database (2001)


```

RESULT 11
XLAJ4126      2012 bp      mRNA      linear      VRT 21-JAN-2000
LOCUS      XLAJ4126      Xenopus laevis erg gene (erg_E).
DEFINITION      AJ224126
ACCESSION      AJ224126.1 GI:5420047
VERSION      ERG gene: transcription factor.
KEYWORDS      Xenopus laevis.
SOURCE      Xenopus laevis.
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
Xenopodinae; Xenopus.
1 (bases 1 to 2012)
Baltzinger, M., Mager-Heckel, A.M. and Remy, P.
X1 erg: expression pattern and overexpression during development
plead for a role in endochelial cell differentiation
Dev. Dyn. 216 (4-5), 420-433 (1999)
JOURNAL MEDLINE 20096678
PUBMED 10633861
REFERENCE 2 (bases 1 to 2012)
AUTHORS      Baltzinger, M.
TITLE      Direct Submission
JOURNAL      Submitted (24-FEB-1998) Baltzinger M., UPR9005, MMDC, Centre
National de la Recherche Scientifique, 15, RUE Rene Descartes,
67084, FRANCE
FEATURES
source      location/Qualifiers
1..2012
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone_lib="lambda ZAPit-erg E"
/dev_stage="stage 30 embryo"
225..1595
/gene="erg"
225..1595
/gene="erg"
/function="transcription factor"
/codon_start=1
/protein_id="CAB46567.1"
/db_xref="GI:5420048"
/translacion="MIOTVDPSSAHIKELSVSEDSLFECTYGNPHLTKTEMTASS
SSDYGOTSKMSPVPODWLSQPSRYTIKMCNPSQVNGSRSPDDCSITGSKKGG
GSDNMGNTGSTEERKHSIPNMTTNERVIVPADFTLSTDIHVRQWLMAKEGLP
DNDVLEQSIDKELCKMTEDFHLRLPSYMAIDILSHLHYLERASFIFFNSVYQ
DANORIPSDLSYEPSRSAMTNHPAPSKASQPTVPTKEDPPOLDPYOILCPT
SSRLANPGGGOIOMQFLLELSDSNKNCITWEGNGEFKMTDPEVARRMERKSK
PMNNTDKLSRALRYDYDKNIMTKVHGKRYAKYEDFGIALQPHPESTMYKPSSEI
PYNSYHAIHPQKMFVAFHPHPALPYTSSSFPAFAPNATYNSPTGSIIPNTRLPASHMS
HLETTY"
BASE COUNT      634 a      479 c      395 g      504 t
ORIGIN
Alignment Scores:
Pred. No.:      2.5e-129      Length:      2012
Score:      2015.50      Matches:      383
Percent Similarity:      81.388      Conservative:      19
Best Local Similarity:      77.538      Mismatches:      23
Query Match:      77.888      Indels:      69
DB:      5      Gaps:      5
US-09-902-772-4 (1-478) x XLAJ4126 (1-2012)
QY      5      ILEYSGLUALALEUSeRValValSerGIUASpQInSerLeuphegIucYsAlaYrGly 24
|||||
Db      258      ATCAAGAGACCTCTTCAGTGGTGAAGAGATCAGTCGTATTTCGAGTGCACCTACGA 317
QY      25      SerProHisLeuAlaYsthrGluMetThrAlaSerSerSerSerGIuTYrGlyGlnThr 44
|||||
Db      318      AGCCCAATCTCTTAAAGAGATGACCGCATCTTCCACAGTCATGTGGCAACCC 377
QY      45      SetLysMetSerProArgValProGInGInAsPTrLeuSerGInProProAlaArgVal 64
|||||

```

```

Db      378      TCGAAATGAGTCACGCTGTGCTCAACAAAGACTGGCTCTCCACAGCCTCCATCCAGGCTG 437
QY      65      ThrIleLysMetGIuCYsAsnProAsnGInValAsnGlySerAlaArgAsnSerProAsp 84
|||||
Db      438      ACCATCAAGATGGAATGCAACCCCTAGTCAATGATGATCAAGAGAGCTCCCGGATGAC 497
QY      85      CysSerValAlaLysGIuGlyLysMetValSerSerSerSerAsnValGlyMetAsnTYr 104
|||||
Db      498      TGCAGTATAGGAAAGAACCAAAATGGTGGAGGGTCGGATATCTTGGAAATCAATTAT 557
QY      105      GlysertyMetGIuGIuLYSHSILeProProAsnMetThrTrAsnGluArg 124
|||||
Db      558      GGCAGCTACATAGAGAAACACATTTCTCCCAACATGACACCAACATGAGAGAGA 617
QY      125      ValIleValProAlaAspProThrLeuTrpSerThrAspHisValAGGInTrpLeuGIu 144
618      GTGATTTACTGCTGACCCCTACTTGTGGAGCAGACATCATGTGCACAGTGGCTAGAG 677
Db      145      TrpAlaValLysGIuTYrGlyLeuProAspValAlaSpIleLeuLeupheGInAsnIleAsp 164
|||||
Db      678      TGGCAATTAAGAAATAGCGGCTTCCAGATGTCATGTGCTTCCCAAAACATGAC 737
QY      165      GLYLsGIuLeuCYsLysMetThrLysAspAspPheGInArgLeuThrProSerTYrAsn 184
|||||
Db      738      GGCAGGATTTATGCAAGATGACCAAGAAAGATTTCATGACACCTACGACCAAGCTACAA 797
QY      185      AlaAspIleLeuLeuSerHisLeuHisTYrLeuArgGlnThrProLeuProHisLeuThr 204
|||||
Db      798      GCTGATATCTCTTCTCGCATCTTCCACTACTCAGAAAGA----- 839
QY      205      SerAspAspValAspLysAlaLeuGInAsnSerProArgLeuMetHisAlaArgAsnThr 224
839-----
QY      225      GLYLAlaThrPheIlePheProAsnThrSerValTYrProGInAlaThrGlnArgIle 244
|||||
Db      840      ---GGTCCCTCTTATTATTTCCCAACTCATCCGTTATACAGCGCAACCAAGATAC 896
QY      245      ThrThrArgProAspLeuProTYrGluGlnAlaArgArgSerAlaTrpThrSerHisSer 264
|||||
Db      897      CCCAGCAGGCAAGATTATCATATGAGCCCTCAAGAGATCTGTTGGACAAATCATCCT 956
QY      265      HisProThrGlnSerLysAlaThrGlnProSerSerSerThrValProLYSThrGlnAsp 284
|||||
Db      957      GCACCA---CCATCAAAAGCATCTCAACCA---TCTACACAGTTCCTCCAAACAGAGAC 1010
QY      285      GlnArgProGlnLeuAspProTYrGlnIleLeuGlyProThrSerSerArgLeuAlaAsn 304
|||||
Db      1011      CCCAGGCCACAGCTAGATCTTATCAGATCTTGGGCCAACAGCACGCACTTGCAAA 1070
QY      305      ProGlySerGIuGlnIleGlnLeuTrpGlnPheLeuLeuGlnLeuLeuSerAspSerSer 324
|||||
Db      1071      CCAAGAGTGGGCGAGATTACAGCTCTGCAATTTCTTATGCAAGCTCTTGGAGATGCTCC 1130
QY      325      AsnSerAsnCYsIleThrTrpGluGlyThrAsnGlyLubPheLysMetThrAspProAsp 344
|||||
Db      1131      AACTCCAACTGCAATCTTACTTGGGAAGCAACCAATGGAATTAAGATGACCGATCTCGAT 1190
QY      345      GluValAlaArgArgTrpGlyGluArgLysSerLysProAsnMetAsnTYrAspLysLeu 364
|||||
Db      1191      GAGGTGGCAAGACGTGGGGGAGAGAGAAAGCAAAACCAACATGAACTATGATTAAGCTC 1250
QY      365      SerArgAlaLeuArgTYrTYrTYrAspLysAsnIleMetThrLysValHisProProGlu 384
|||||
Db      1251      AGCGGTGCACTTGTTACTACTCTAGATGAATAAATATATATGACTAAAGTC----- 1298
QY      385      SerSerMetLYrLysTYrProSerAspLeuProTYrMetSerSerTYrHisGlyLysArg 404
|||||
Db      1299-----
QY      405      TYrAlaTYrLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnPro----- 420
|||||
Db      1311      TATGCTTACAAATTTGATTTTTCATGGGATGCTCAAGCTCTCAACCTCATCTCCAGAA 1370

```

QY 421 -----HisAlaHisPro 424
DB 1371 TCACTATGTACAGATGCTCCAGAACTTCCATATATGAGTCATACATGACATGACATGCA 1430
QY 425 GlnlyMetAspPheValAlaProHisProProAlaLeuProValThrSerSerPhe 444
DB 1431 CAGAAAGATGATGTTGTTGCCCCCATCCGCGCATCTTACGTGATCATCTCAAGTTTC 1490
QY 445 PheAlaAlaProAspProTyrThrPheSerProThrGlyGlyIleTyrProAsnThrArg 464
DB 1491 TTTCGACGCCCCCTAATGCAATGCGAATTCACCACTGGAAGATTTTATTCAAATACTCGG 1550
QY 465 LeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr 478
DB 1551 CTGCGACGTCAGCCATATGCTTCTCATCTTGGAACTTACTAT 1592
RESULT 12
HOMERG11
LOCUS HOMERG11 3126 bp mRNA linear PRI 27-APR-1993
DEFINITION Human erg protein (ets-related gene) mRNA, complete cds.
ACCESSION M21535 M17390
VERSION M21535.1 GI:182182
KEYWORDS erg protein.
SEGMENT 1 of 2
SOURCE Human, cell line COLO 320, cDNA to mRNA, lambda-7.
ORGANISM Homo sapiens
Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
AUTHORS Reddy, E.S., Rao, V.N. and Papas, T.S.
TITLE The erg gene: a human gene related to the ets oncogene
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 84 (117), 6131-6135 (1987)
MEDLINE 87317608
PUBMED 3476934
FEATURES
source Location/Qualifiers
1..3126
/organism="Homo sapiens"
/db_xref="taxon:9606"
195..1286
/note="erg1 protein"
/codon_start=1
/protein_id="AA35811.1"
/translation="MVGSDPTVGMNYSYMEKMKMPNNMTTNERVITVPADPTLMST
DHVQMLMAVKEGLPDVNIILFQNIQKELCKKDFORLTPSYNADILSLHY
LREPLPHLTSDVDVGLALONSPRLMHANRDLPPPRSAWGTHGHPPOSKAOPS
PSTVKTEDORPOLDPYOLIGPTSSRLANPSGOTLOLQFLEILSDSSNSCTWEG
TNGEKMTDPPEVARRMSEKSKPMNTDKLSRLRYTDKIKMTKVKGRKRYKEDF
HGIALDLPHPPESSLYKTPSDLPYMGSTYHAPQAMNFVAHPHPALPYTSSSEFAAPN
PYWNSPTGILPNTRLPRTSHMPSHLGTY"
BASE COUNT 928 a 732 c 725 g 741 t
ORIGIN 1 bp upstream from EcoRI site.
Alignment Scores:
Pred. No.: 6,896-115 Length: 3126
Score: 1808.50 Matches: 342
Percent Similarity: 82.23% Conservative: 5
Best Local Similarity: 81.04% Mismatches: 10
Query Match: 69.88% Indels: 65
DB: 9 Gaps: 4
US-09-902-772-4 (1-478) x HOMERG11 (1-3126)
QY 78 SerArgAsnSerProAspAspCysSerValAlaIleGlyMetValSerSer 97
DB 150 TCAAGGAAGCTCTGATGATGATGAGTGGCCAAAGCGGGAAGATGGTGGCCACCA 209
QY 98 AspAsnValIleGlyMetAsnTyrGlySerTyrMetGluGluHisIleProProAsn 117
DB 210 GACACCGTTGGGATGACACGCGACCTACATGAGAGAAACACACATGCCACCCCAAC 269
QY 118 MetThrThrAsnGluArgValIleValProAlaAspProThrLeuTyrSerThrAsp 137

DB 270 ATGACCACGAGAGCGAGAGTATGCTGCGCAGAGATCCAGCTATGAGGTACAGAC 329
QY 138 HisValArgGluThrPheGluIleThrPheValIleGlyTyrGlyLeuProAspValAspIle 157
DB 330 CATGTGCGGCAATGCTGAGAGTGGGCGGTGAAGAATATGGCTTCCAACTCAACATC 389
QY 158 LeuLeuPheGluAsnIleAspGlyLysGluLeuCysLysMetThrLysAspPheGlu 177
DB 390 TTGTTATTCACAAACATTCATGAGGAAGAACTGTGCAAGATGACCAAGACACTCCAG 449
QY 178 ArgLeuThrProSerTyrTrpAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgIu 197
DB 450 AGGCTACCCCGACGATACAGCCGACATCTTCTCTCATCTCAGTCTCAGTACCTCAGAGAG 509
QY 198 ThrProLeuProHisLeuThrSerAspAspValAspIleAlaLeuGlnAsnSerProArg 217
DB 510 ACTCCCTTCCACATTTGACTTCAGATGATGTGTAATAAGCCTTACAAACCTCCACGG 569
QY 218 LeuMetHisAlaArgAsnThrGlyAlaThrPheIlePheProAsnThrSerValTyr 237
DB 570 TTAATGCATGTAGAAACACA----- 590
QY 238 ProGluAlaThrGluArgIleThrThrArgProAspLeuProTyrGluIleAlaArgArg 257
DB 591 -----GATTATACATATGAGCCCGCCAGAGAG 617
QY 258 SerAlaThrThrSerHisSerHisProThr---GlnSerLysAlaThrGlnProSerSer 276
DB 618 TCAAGCTGACCGGTCACGCGCCACCCAGCCCGCCAGTCCAAAGCTGCTCAACCATCTCT 677
QY 277 SerThrValProLysThrGluAspGlnArgProGlnLeuAspProTyrGlnIleLeuGly 296
DB 678 TCCACAGTCCCAAAACGTAAGACAGCGCTCCAGTTGATGATCATTACAGATTCTTGGA 737
QY 297 ProThrSerSerArgLeuAlaAsnProGlySerGlyGlnIleGlnLeuThrPheLeu 316
DB 738 CCACACAGTAGCGGCTTCCAAATCCAGGACATGCGCATCTTGGCAGTTCCTC 797
QY 317 LeuGluLeuLeuSerAspSerSerAsnSerAsnGlyIleThrTrpGluIleThrAsnGly 336
DB 798 CTGGAGACTCTGTGCGACGCTCCAACTCCAGCTGATCATCTGGAGAGCCCAACGGG 857
QY 337 GluPheLysMetThrAspProAspGluValAlaArgArgTyrGlyLysGlySerLys 356
DB 858 GAGTTCAAGATGACGATCCGACGACGAGGCGCCGCTGGGAGAGCGGAAGACAA 917
QY 357 ProAsnMetAsnTyrAspLysLeuSerArgAlaLeuArgTyrTyrTrpAspLysAsnIle 376
DB 918 CCCACATGACATGATGATGATGCTCAGCCGCGCTCCGTTACTACTATGACAAGAACATC 977
QY 377 MetThrLysValHisProProGluSerSerMetTyrLysTyrProSerAspLeuProTyr 396
DB 978 ATGACCAAGTC----- 989
QY 397 MetSerSerTyrHisGlyLysArgTyrAlaTyrLysPheAspPheHisGlyIleAlaGln 416
DB 990 -----CARGGAAGCGGTACGCGCTCAAGTTGAGCTCCACGCGATGCGCCAG 1037
QY 417 AlaLeuGlnPro----- 420
DB 1038 GCCCTTCACCCCGCCCGGAGTCAATCTGTACAGTACCCCTCAGACCTCCCGTAC 1097
QY 421 -----HisAlaHisProGlnLysMetAsnPheValAlaProHisProProAla 436
DB 1098 ATGGGCTCTTATCAGCCCAACACAGAGATGAATTTGTGGCGCCCACTCCAGGC 1157
QY 437 LeuProValThrSerSerSerPhePheAlaAlaProAsnProTyrTrpAsnSerProThr 456
DB 1158 CTCCCGGTACACATCTTCAGTTTGTGTCGCCCAAAACCCATACGGAATTCACCAACT 1217
QY 457 GlyGlyIleTyrProAsnThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThr 476

Db 1218 GGGGTAATATACCCCACTAGGCTCCGCCACCAACCATATGCTTCTCATCTGGGCACT 1277
QY 477 TyrTyr 478
Db 1278 TACTAC 1283

RESULT 13
DRE249590
LOCUS DRE249590 1436 bp mRNA linear VRT 03-MAR-2000
DEFINITION Danto rerio mRNA for Fil-1 protein (fil-1 gene).
ACCESSION AJ249590
VERSION AJ249590.1 GI:6006476
KEYWORDS fil-1 gene; Fil-1 protein.
SOURCE Danto rerio.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danto.

REFERENCE 1 (bases 1 to 1436)
AUTHORS Brown,L.A., Rodaway,A.R., Schilling,T.F., Jowett,T., Ingham,P.W., Patient,R.K. and Sharrocks,A.D.
TITLE Insights into early vasculogenesis revealed by expression of the ETS-domain transcription factor Fil-1 in wild-type and mutant zebrafish embryos
JOURNAL Mech. Dev. 90 (2), 237-252 (2000)
MEDLINE 20108585
PUBMED 10640707

REFERENCE 2 (bases 1 to 1436)
AUTHORS Sharrocks,A.D.
TITLE Direct Submission
SUBMITTED (29-SEP-1999) Sharrocks A.D., School of Biological Sciences, University of Manchester, Oxford Road, Manchester, M13, 9PT, UNITED KINGDOM

FEATURES
source
1. 1436
/organism="Danto rerio"
/db_xref="taxon:7955"
/dev_stage="20-24hr embryo"
45..1400
/gene="fil-1"
45..1400
/gene="fil-1"
/function="transcription factor"
/codon_start=1
/evidence="experimental"
/product="Fil-1 protein"
/protein_id="CAB56832.1"
/db_xref="GI:6006477"
/translation="MDGTLKEALSVSEDSQSLPEPPYAAAPLPKTDMTASGTODYGQ
THKINIPQOEMINQPVAVNVKREYDHLNGSRSESPVDCSKCNKMGVGTASOMNY
TGKMDKCAPPMNTNERVIVPADPSIDMSPDHYROMDIAIKEYGLQIDTAMFHS
TGKELCKSKDDELRLTSVNTVEVLSHLNLTRESSISYNTPEHAOSPLAKND
DASYDAVRRGTGSMNNHSGKSPYVSVOSKRNPDQPRQPOYOTIGPTSSRLNPG
SGOILUMOFILFLSDSANAGCITWGTGGERKMDPDVAVRMGRSKPKMNTDKL
SRALRTYDKNIMTKVHKRYAKFDFHGIQALQHPTESTMTKIPSELVYPVSTHA
HOOKVNFSPHPSPMPTVSSNFPETTPWSSPPGCIYPPSVPRHANSVSHLSGY
Y"

gene
CDS
BASE COUNT 377 a 439 c 349 g 271 t
ORIGIN

Alignment Scores:
Pred. No.: 1.52e-93 Length: 1436
Score: 1491.50 Matches: 289
Percent Similarity: 68.13% Conservative: 53
Best Local Similarity: 57.57% Mismatches: 85
Query Match: 57.63% Indels: 75
DB: 5 Gaps: 7

US-09-902-772-4 (1-478) x DRE249590 (1-1436)

QY 1 MetalaserThrIlelysgIuaIaleuSerValValserGIuaspGIinsertleupheGlu 20
Db 45 ATGAGCAGGAACATTAAGAGGCGCCTGTCTCAGTGTAGTGAAGACCAAGTCTCTGTTCGAG 104

QY 21 CysAlaTyr---GlySerProHisIleuAlaIalysThrGluMetThrAlaSerSerSer 39
Db 105 CCTCATACAGCGCGCTGCTGCTCTTACCCAGACAGACATGACTGGCTCGGACATCAG 164

QY 40 GluTyrGlyGlnThrSerIlyMetSerProArgValProGlnGlnAspTyrPleuSerGln 59
Db 165 GATTACGGCCAGACGCAAAATCAACCCATCCCTCCAGCAAGATGATTAACAG 224

QY 60 ProProAlaArgValThrIlelyMetGluCysasnProasnGlnValasnGlySerArg 79
Db 225 ---CCGGTGGGGGTGAACGTCAACGAGACTAT-----GACCACATCAACGGATCCAGA 275

QY 80 AsnSerProAspAspCysSerValAlaIalysGlyIlyMetValSerSerSerAspAsn 99
Db 276 GAGTCGGCGGTGAGCTCAGCTCAGCTGGGAAATGCAATTAATGGTGGGGGAGCGAAGCG 335

QY 100 ValGlyMetAsnTyrGlySerTyrMetGluGluLysHisIleProProAsnMetThr 119
Db 336 TCTCAGATGAACATATAGTGGCTACATGATGAGAAATGTGCTCCGCTCCAAACATGACG 395

QY 120 ThrAsnGluArgArgValIleValProAlaAspProThrIleuTyrPleuSerThrAspHisVal 139
Db 396 ACCAATGAGAGAGATCATCTGCCGACAGCCGTCGTGTGTCTCCGGATCAGCTG 455

QY 140 ArgGlnTyrPleuGluTyrAlaValIalysGlyIlyTyrGlyLeuProAspValAspIleLeuLeu 159
Db 456 CGGAGTGTAGACGTGCGGCATTAAGAGTACGCTTTCAGAGATCGACAGCGGCATG 515

QY 160 PheGlnAsnIleAspGlyIlyGluLeuCysIlyMetThrIlyAspAspPheGlnArgLeu 179
Db 516 TTTCACAGCACAGATGGAAAGAGCTCTGCAAGATGAGCAAAAGCAGCTTCTGAGACTC 575

QY 180 ThrProSerTyrAsnAlaAspIleLeuSerHisIleuHisTyrLeuArgGluThrPro 199
Db 576 ACCAGCCTTATTAACACCGAGGTCTCTCTCAGATCTCAATTAACCTCAGGAAAGT--- 632

QY 200 LeuProHisIleuThrSerAspAspValAspLysAlaLeuGlnAsnSerProArgLeuMet 219
Db 632 ----- 632

QY 220 HisAlaArgAsnThrGlyGlyAlaThrPheIlePheProAsnThrSerValTyrProGlu 239
Db 633 -----AGTCAATCATATCATACACAGCCCATCTCAGCGTACG 671

QY 240 AlaThrGlnArgIleThrThrArgProAspLeuProTyrGlnAlaArgArgSerAla 259
Db 672 CAGTCCCTCGCTGCGCTGCAAAAGACGATGCGTTATGATGCTGTACGAGAGCGGC 731

QY 260 TrpThrSerHisSerHisProThrGlnSerLysAlaThrGlnProSerSerThrVal 279
Db 732 TGGTCGAACAACATGCACACGAGAAAGGCTCTCCACACATGGCTCTCAGAGCGTCC 791

QY 280 ProLysThrGluAspGlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSer 299
Db 792 AAAAACCAGGATGAGCTCTCTCCACGACGATCTTATGATATGAGACCCACGACG 851

QY 300 SerArgIleuAlaAsnProGlySerGlyGlnIleGlnLeuTrpGlnPheLeuLeuGluLeu 319
Db 852 AGTCAGCTGTCTAACCAGGCTCAGTCAGATCAGATCAGTGTGCGATCTGCGAGACCTC 911

QY 320 LeuSerAspSerSerAsnSerAsnCysIleThrTrpGluGlyThrAsnGlyGluPheLys 339
Db 912 CTGTCCGACAGCGCAACGCGGCTGATATCCTCGGGAAGAACCAAGGAGTTCAAG 971

QY 340 MetThrAspProAspGluValAlaIalysArgTyrGlyGluArgLysSerLysProAsnMet 359
Db 972 ATGACGACCCACAGACGAGTGTGACGCGCTGGGGCGCAAGCAACCAACCATG 1031

QY 360 AsnTyrAspLysLeuSerArgAlaLeuArgTyrTyrTyrAspLysAsnIleMetThrLys 379
Db 1032 AACTACGACAAAGCTGAGCCGAGCCCTGCGTACTACTACGACAAAGACATCATGACCAA 1091

QY 380 ValHisProGluSerSerMetYrLysYrProSerAspLeuProTyrMetSerSer 399
 Db 1092 GTG----- 1094
 QY 400 TyrHisGlyAspArgTyrAlaTyrLysPheAspPheHisGlyLeuAlaGln 419
 Db 1095 ---CACGGACAGCGCTACGCTTACAGTTCGACTTCACGGATGCGCCAGGCCCTGCAN 1151
 QY 420 Pro----- 420
 Db 1152 CCCCATGCCAGCGAGTCACATGTACAGTACCCAGGAGCTGCGCTGCTAGC 1211
 QY 421 ---HisAlaHisProGluLysMetAsnProValAlaProHisProProAlaLeuProVal 439
 Db 1212 TATCCAGGCGCCAGCAGAGGTCAACCTGCTCTCTCCACCTCCGCTCCAGTCCCGC 1271
 QY 440 ThrSerSerSerPhePheAlaAlaProAsnProTyrTyrPheSerProThrGlyGlyLe 459
 Db 1272 ACCTCTCAACTCTTCTGGTCCACACCCCATCTGAGCTCCCGACAGAGGATATT 1331
 QY 460 TyrProAsnThrArgLeuProAla-----AlaHisMetProSerHisLeuGlyThr 476
 Db 1332 TACCCCAATCCAGCGTCCACGCGCATTCACAGCTGCTTACACACCTTGCTAGT 1391
 QY 477 TyrTyr 478
 Db 1392 TACTAC 1397
 RESULT 14
 LOCUS AY029368 1359 bp mRNA linear PRI 09-MAY-2001
 DEFINITION Homo sapiens friend leukemia integration 1 transcription factor
 (FLI1) mRNA, complete cds.
 ACCESSION AY029368
 VERSION AY029368.1 GI:14017402
 KEYWORDS
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 1359)
 AUTHORS Uhl, B.T.S., Ralney, D.R. and Meredith, D.M.
 TITLE Direct Submission
 JOURNAL Submitted (09-APR-2001) Molecular Medicine Unit, Leeds University,
 Beckett St, Leeds LS9 7TF, United Kingdom
 FEATURES
 source
 1..1359
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="11"
 1..1359
 /gene="FLI1"
 1..1359
 /note="RTS transcription factor; involved in the
 pathogenesis of Ewings sarcoma following translocation
 with EWS gene on chromosome 22; may have oncogenic
 potential when fused with EWS gene"
 /codon_start=1
 /product="friend leukemia integration 1 transcription
 factor"
 /protein_id="AAK50443.1"
 /db_xref="GI:14017403"
 /translation="MDGRIKELASVSDQSLPDSAYGAALHPRADWTAGSPDYQ
 PKHINLPDQEWINOPVAVKREYDHMSRESVSCSKLWGGESNPMY
 NSYDEKGPPTMTTERVIVPADTLTQEHVRLMAYIKESLMDISFQ
 NMDGELCKMKEDLRATLTLYNTEVLISHSYRESLKYNTSHPDOSRLSYK
 DPSTDSVRKRWGNMNSGLKSPPLGSAQTISKNTEDRPDPYQILGPTSSRLANP
 GSGQIQIWLFLLELSDSANSCITWETNEEFKMTDDEVAARMGERSKPMNNDK
 LSRALRYYNINIMTKVGRYAKFDFHGTALQLOPPTDESSMKYPSDISYMPSTYH
 AHQGVNVPRPSSMPYTSSTSSFGAASQWYTSFGITGYPNPVRHNPHTVPSHLGS
 YH"
 BASE COUNT 359 a 422 c 333 g 245 t

ORIGIN
 Alignment Scores:
 Pred. No.: 6,36e-93 Length: 1359
 Score: 1482.00 Matches: 291
 Percent Similarity: 68.65% Conservative: 55
 Best Local Similarity: 57.74% Mismatches: 80
 Query Match: 57.26% Indels: 78
 Ds: 9 Gaps: 9
 US-09-902-772-4 (1-478) x AY029368 (1-1359)
 QY 1 MetAlaSerThrIleLysGlnAlaLeuSerValAlaSerGlnAspGlnSerLeuPheGlu 20
 Db 1 ATGAGCGGGCATTAAGAGAGCTGTGTGGTGGAGGACGACAGCCCTCTTTGTAC 60
 QY 21 CysAlaTyrGly---SerProHisLeuAlaLysThrGlnMetThrAlaSerSerSer 39
 Db 61 TCAGCGTACGAGCGGCGACCCATCTCCCAAGGCGGACATAGTCCCTCGGGAGTCT 120
 QY 40 GluTyrGlyGlnThrSerLysMetSerProArgValProGlnGlnAspThrLeuSerGln 59
 Db 121 GACTACGGGACAGCCCAAGATCAACCCCTCCACAGACAGAGTGTATCATCG 180
 QY 60 ProProAlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArg 79
 Db 181 ---CCAGTGGGGTCAACGCTCAAGCGGAGTAT-----GACCACATGAATGATCCAG 231
 QY 80 AsnSerProAspAspCysSerValAlaLysGlyLysMetValSerSerSerAspAsn 99
 Db 232 GAGTCTCCGGTGAGTGCAGCGCTTACCAATGTCAGCAAGCTGTGGCGGAGCGATCC 291
 QY 100 ValGlyMetAsnTyrGlySerTyrMetGluGluLysHis---IleProProAsnMet 118
 Db 292 AACCCCATGAATACAAACAGTATATGACGAGAGAAGATGCCCTCCCTCCCAACATG 351
 QY 119 ThrThrAsnGluArgArgValIleValProAlaAspProThrLeuThrSerThrAspHis 138
 Db 352 ACCACCAAGAGAGAGAGTATCGTCCCGACAGCCCACTGTGGCAGACAGAT 411
 QY 139 ValArgGlnThrLeuGluThrValAlaValLysGluTyrGlyLeuProAspValAspIleLeu 158
 Db 412 GTGAGCGAATGGCTGAGTGGGCGCATTAAGAGTACAGCTTATGAGATGCACATCC 471
 QY 159 LeuPheGlnAsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArg 178
 Db 472 TTTTCCAGAACATGATGATGCAAGGAACTGTAAATGACAAAGAGAGACTCTCCGCG 531
 QY 179 LeuThrProSerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGluThr 198
 Db 532 GCCACCACTCTACAAACAGGAGAGTGTGTACACCTCAGTACCTCAGGGAAGT 591
 QY 199 ProLeuProHisLeuThrSerAspValAspLysAlaLeuGlnAsnSerProArgLeu 218
 Db 592 TCAGTG----- 597
 QY 219 MetHisAlaArgAsnThrGlyAlaThrPheIlePheProAsnThrSerValTyrPro 238
 Db 598 -----CTGGCTATAAATCAACCTCCACACC 624
 QY 239 GluAlaThrGlnArgIleThrThrArgProAspLeuProTyrGluGlnAlaArgArgSer 258
 Db 625 GACCAATCCCTCAGATGCTGTGTCAAAGAACCTCTTAACTCATGTCAGGAAGAGA 684
 QY 259 AlaThrPheSerHisSerHisProThrGlnSerLysAlaThrGln---ProSerSerSer 277
 Db 685 GTTGGGGCAATAAATGAATCTGGCTCAACAAAGCTCCCTTGAGAGGGACACA 744
 QY 278 ThrValProLysThrGlnAspGlnArgProGlnLeuAspProTyrGlnIleLeuGlyPro 297
 Db 745 ACGATCAGTAAGAAATACAGACGACGAGCCGACAGATCCGTATCAGATCTCGGGCCG 804
 QY 298 ThrSerThrGluLeuAlaAsnProGlySerGlyGlnIleGlnLeuThrGlnPheLeuLeu 317


```
Db 477 ACCACCAAGAGAGAGATGATGTCCTCCCGACACCCACAGCTGAGACACAGGACAT 536
QY 139 ValArgGlnTrpLeuGlnTrpAlaValLysGluTyrGlnLeuProAspValAspLeu 158
537 GTGAGGCAATGGCTGGAGTGGCCATTAAGAGATACAGCTTGTGAGATCGACACATCC 596
QY 159 LeuPheGlnAsnLysPglLysGluLeuCysLysMetThrLysAspPheGlnArg 178
597 TTTTTCAGAACATGATGGTGAAGAACTGTGTAATAAGCAAGAGGACTTCCTCCGC 656
QY 179 LeuThrProSerTyrAsnAlaAspLeuLeuSerHisLeuHisTyrLeuArgGluThr 198
657 GCCACCAACCTCTACAAACAGGAGAGTCTTGTCTACACCTCAGTTACTCAGGAAAGT 716
QY 199 ProLeuProHisLeuThrSerAspValAspLysAlaLeuGlnAsnSerProArgLeu 218
717 TCACTG----- 722
QY 219 MethHisAlaArgAsnThrGlnGlyAlaThrPheLeuPheProAsnThrSerValTyrPro 238
723 -----CTGGCTATATACACCTCCACACC 749
QY 239 GluAlaThrGlnArgGlnThrThrArgProAspLeuProTyrGlnAlaArgArgSer 258
750 GACCAATCCTCAGATGATGATGTCAAAGAGACCCCTTGTATGACTCAGTCAGAGAGAGA 809
QY 259 AlaTrpThrSerHisSerHisProThrGlnSerLysAlaThrGln--ProSerSerSer 277
810 GCTTGGGGCAATGAATGATTCGTGGCTTCACAAAAGTCTCCCTTGAGAGGGGCACAA 869
QY 278 ThrValProLysThrGlnAspGlnArgProGlnLeuAspProTyrGlnLleuGlyPro 297
870 ACATGATGATGAATACAGAGCAACGCCCCACAGATCCGTATCAGATCTGGGGCCG 929
QY 298 ThrSerArgGluAlaAsnProGlnLysSerGlyGlnLleuLeuTrpGlnPheLeuLeu 317
930 ACCAGCAGTGCCTACCAACCTCGAAGGGGAGATCCAGCTGTGGCAATTCCTCTG 989
QY 318 GluLeuLeuSerAspSerSerAsnSerAsnCysIleThrTrpGluGlyThrAsnGlyGlu 337
990 GAGCTGCTCTCCGACAGCGCAACGCGACGCTGTATCAGCTGGAGGGAGCAACGGGAG 1049
QY 338 PheLysMetThrAspProAspGluValAlaArgTrpGlnGluArgLysSerLysPro 357
1050 TTCAAATATGAGGAGCCCGATGAGTGGCCAGCGCTGGGGCGAGGAAAGCAAGCCC 1109
QY 358 AsnMetAsnTyrAspLysLeuSerArgAlaLeuAlaArgTyrTyrAspLysAsnLleMet 377
1110 AACATGAATTACGACAAGCTGAGCCGGCCCTCCGTTATTACTATGATATAAAACATTATG 1169
QY 378 ThrLysValHisProGlnLeuSerMetTyrLysTyrProSerAspLeuProTyrMet 397
1170 ACCAAGTG----- 1178
QY 398 SerSerTyrHisGlyLysArgTyrAlaTyrLysPheAspPheHisGlyLleAlaGlnAla 417
1179 -----CACGGCAAAAGATATGCTTACAAATTGACTTCCACGCGCATTTGCCAGGCT 1229
QY 418 LeuGlnPro----- 420
1230 CTGCAGCCACATCCGACCGAGTGTCCATGTACAACTACCCCTTGACATCTCCTACATG 1289
QY 421 -----HisAlaHisProGlnLysMetAsnPheValAlaProHisProAlaLeu 437
1290 CTTCTACCATGCCACAGGAGGTGAATTTGTCCCTCCCATCATCATCTCCATG 1349
QY 438 ProValThrSerSerSerPhePheAlaAlaProAsnProTyrTrpAsnSerProThrGly 457
1350 CCGTACTCTCCCTCAGCTTCTTGGAGCGCATACAAATACGACCTCCCGCCAGGGG 1409
QY 458 GlyLleTyrProAsn-----ThrArgLeuProAlaAlaHisMetProSerHisLeu 474
|||||
```

```
Db 1410 GGAATCTACCCCAACCCCAAGTCCCGCCCATCTTACACCCACAGTGCCTTCACTTA 1469
QY 475 GlyThrTyrTyr 478
Db 1470 GGCAGCTACTAC 1481
```

Search completed: November 9, 2002, 18:55:34
Job time : 3121.42 secs

FT /+tag- a
XX JP11075871-A.
XX
PD 23-MAR-1999.
XX
PF 29-MAY-1998; 98BP-0166076.
XX
XX 20-JUN-1997; 97US-0050297.
PR 18-JUN-1997; 97US-0878177.
XX
PA (CHUGAI PHARM CO LTD.
PA (UYPE-) UNIV PENNSYLVANIA.
XX
DR WPI; 1999-257708/22.
DR P-PSDB; AA01521.
XX
PT An active protein for inhibiting cell calcification - useful for
PT measuring the calcification of a cell, for diagnosing arthritis
PT deformans or ossification of spinal column ligament
XX
PS Disclosure; Page 8-9; 15pp; Japanese.
XX
CC The present sequence encodes a chicken c-ery protein. The specification
CC also describes a chicken C-11 protein (AA01520) which has cell
CC calcification inhibiting activity and a cell calcification inhibiting
CC agent containing c-ery protein. The proteins are used for measuring the
CC calcification of a cell, for diagnosing arthritis deformans or
CC ossification of spinal column ligament.
XX
SO Sequence 1528 BP; 464 A; 395 C; 330 G; 339 T; 0 other;

Alignment Scores:
Pred. No.: 1,95e-192 Length: 1528
Score: 2436.00 Matches: 458
Percent Similarity: 91.97% Conservative: 0
Best Local Similarity: 91.97% Mismatches: 0
Query Match: 94.13% Indels: 40
DB: 20 Gaps: 2

US-09-902-772-4 (1-478) x AA026552 (1-1528)
QY 1 MetAlaSerThrIleuysGluAlaIeuSerValSerGluAspGlnSerIleuPheGlu 20
DB 63 ATGGCAAGGCACTATTAAAGGACATTATCAGTGTAGTACAGACCAAGCTCTGTTGAG 122
QY 21 CysAlaTyrlGlySerProHlsIeuAlaIysThgluMetHraIaSerSerSerGlu 40
DB 123 TGTGCTACGGATCGCCCACTTGCAGAGACAGAAATGACAGCTCTCTTCAGTGAA 182
QY 41 TyrGlyGlnThrSerIysMetSerProAlaGValProGlnGlnAspTyrIleuSerGlnPro 60
DB 183 TATGGGCAAAACATCAAGATGAGCCGCGCTTCCAGAGACTGTGTATACAGCCC 242
QY 61 ProAlaArgValThrIleuysMetGluCysAsnProAsnGlnValAsnGlySerArgAsn 80
DB 243 CCGGCCAGACTTCCATTAAAGATGAGTGTAAACCAACAGGTAAATGGGTCAAGGAAT 302
QY 81 SerProAspAspCysSerValAlaIalysGlyIlyuMetValSerSerSerAspAsnVal 100
DB 303 TCACGTGATGATCGACGCGGCAAAAGGGAATATGTTAGCAGTTACAGCAATGTT 362
QY 101 GlyMetAsnTyrlGlySerTyrMetGluGluIlyuHisIleProProAsnMetThrThr 120
DB 363 GGGATTAACATATGGAAGCTCATGAGAAAGAAACATATTCCTCCAAATATGACAAAC 422
QY 121 AsnGluArgArgValIleValProAlaAspProThrIleuThrSerThrAspHisValArg 140
DB 423 AATGAACGAAAGATTTATGTCAGAGATCTACGTTAAGGACACAGCCATGTAAGG 482
QY 141 GlnTyrIleuGlnTyrPalaValIysGluTyrGlyIleuProAspValAspIleuLeuPhe 160
DB 483 CAGTGGCTGGAGTGGGACAGTGAAGAGATATGCTTCACAGACGTGACATCTTGTGTTTC 542

QY 161 GlnAsnIleAspGlyIysGluIleuCysIysMetThrIysAspAspPheGlnArgLeuThr 180
DB 543 CAGAAACATGTGAGGAAGAGTTGTAAATGACCAAGATGACTTCCAGAGACTCAGC 602
QY 181 ProSerTyrAsnAlaAspIleuLeuSerHisIleuHisIleuArgGluThrProLeu 200
DB 603 CCGAGCTATTAACGCAAGATATCTCTGTACACCTACCTACCTACAGAGAGACTCTCT 662
QY 201 ProHisLeuThrSerAspAspValAspIysAlaIleuGlnAsnSerProAlaGlyMetHis 220
DB 663 CCACATTTGACTTCAGATGATGTTGATAGGCCCTTACAAAACCTCTCCACGGTTATGAT 722
QY 221 AlaArgAsnThrGlyIlyalThrPheIlePheProAsnThrSerValTyrProGluAla 240
DB 723 GCTAGAAACACAGAGAGGACACTTATTTTCCAAATGATCATCAGTTACCCAGAAAGA 782
QY 241 ThrGlnArgIleThrThrArgProAspLeuProTyrGluIleuAlaArgAspSerIleu 260
DB 783 ACGCAAGAAATTAACAAACAGGCGCAGATTTACTTATGAGCAAGCAGAGATCAGCCTGG 842
QY 261 ThrSerHisSerHisProThrGlnSerIysAlaThrGlnProSerSerThrValPro 280
DB 843 ACGAGTACAGCCATCCCACTCAGTCAAAAGCTAACCAACATCATCTTCAACAGTCCC 902
QY 281 LysThrGluAspGlnArgProGlnIleuAspProTyrGlnIleuGlyProThrSerSer 300
DB 903 AAAACAGAAACACAGCGCTCCTCAGTTGATCTTATCAGATCTTGAGACGACAGCAGC 962
QY 301 ArgLeuAlaAsnProGlySerGlyGlnIleGlnLeuThrPheIleuLeuGluLeu 320
DB 963 CGTCTTGCAATCCAGGAGGATGGCAGATACCTTACCTAGGAGTTCCTACCTGAGCTTCTG 1022
QY 321 SerAspSerSerAsnSerAsnCysIleThrTyrGluGlyThrAsnGlyIlyuPheIysMet 340
DB 1023 TCGGACAGCGTCCAACTCCACTGATCAGTCTGGAGGCGCAAAATGGGAGTTCAAGATG 1082
QY 341 ThrAspProAspGluValAlaArgArgTyrGlyIlyuArgIysSerIysProAsnMetAsn 360
DB 1083 ACAGACCTGATGAAGAGCGCTCGGCGTGGGAGAGAGAAAGCAAACTACATGAGAAC 1142
QY 361 TyrAspIysLeuSerArgAlaIleuArgTyrTyrTyrAspIysAsnIleuMetThrIysVal 380
DB 1143 TATGACAAACTCAGCGGTGACTTCTGCTACTATGACAAAAATATATGACTAAAGTT 1202
QY 380 ----- 380
DB 1203 CATGTAAACGCTATGCTCAAAATTTGATTTCCACAGGAATCGCTCAGGCCCTCCAGCCT 1262
QY 381 HisProProGluSerSerMetTyrIysTyrProSerAspLeuProTyrMetSerSerTyr 400
DB 1263 CACCTTCCAAAATCAATCAATGATCAAAATACCATCAGACCTCCCTTACATGATCTCTC 1322
QY 401 HisGlyIysArgTyrAlaTyrIysPheAspPheHisGlyIleAlaGlnAlaIleuGlnPro 420
DB 1322 ----- 1322
QY 421 HisAlaHisProGlnIlyuysMetAsnPheValAlaProHisProProAlaIleuProValThr 440
DB 1323 CATGCACACCCCAAGAAAGATGAACTTTGATGCTCCCATCCCTCCCTGCTTGGCCGTAAC 1382
QY 441 SerSerSerPhePheAlaIleProAsnProTyrTyrAsnSerProThrGlyIlyIleTyr 460
DB 1383 TCATCCAGCTTTTCTGCGCCCTAATCACTAGTGAATTCACCAACTGGAGGACATCTAC 1442
QY 461 ProAsnThrArgLeuProAlaAlaHisMetProSerHisIleuGlyIlyThrTyrTyr 478
DB 1443 CCCAATTAACAGGCTGCGACGTCTCATATGCTTCCATCTTGGACCTACTAC 1496
RESULT 2
AA026551
ID AA026551 standard; DNA; 1447 BP.
XX

AC AAX26551;
 XX 14-JUN-1999 (first entry)
 DE DNA encoding chicken C-11 protein.
 XX Chicken; C-11 protein; cell calcification inhibiting activity;
 KM cell calcification inhibiting agent; c-ery protein; arthritis deformans;
 KW ossification; spinal column ligament; ss.
 XX Gallus sp.
 OS
 FH Key Location/Qualifiers
 FT CDS 63..1418
 FT /*tag= a
 XX JP11075871-A.
 XX 23-MAR-1999.
 XX 29-MAY-1998; 98JP-0166076.
 XX 20-JUN-1997; 97US-0050297.
 XX 18-JUN-1997; 97US-0878177.
 XX (CHUS) CHUGAI PHARM CO LTD.
 XX (UYPE-) UNIV PENNSYLVANIA.
 XX WPI: 1999-257708/22.
 XX P-PSDB; AAY01520.
 XX
 PT An active protein for inhibiting cell calcification - useful for
 PT measuring the calcification of a cell, for diagnosing arthritis
 PT deformans or ossification of spinal column ligament
 XX
 PS Disclosure; Page 7-8; 15pp; Japanese.
 CC The present sequence encodes a chicken C-11 protein which has cell
 CC calcification inhibiting activity. The specification also describes
 CC a cell calcification inhibiting agent containing c-ery protein
 CC (AAY01521). The proteins are used for measuring the calcification of a
 CC cell, for diagnosing arthritis deformans or ossification of spinal column
 CC ligament.
 XX
 SQ Sequence 1447 BP; 440 A; 374 C; 317 G; 316 T; 0 other;
 Alignment Scores:
 Pred. No.: 4,38e-178 Length: 1447
 Score: 2262.50 Matches: 430
 Percent Similarity: 86.35% Conservative: 0
 Best Local Similarity: 86.35% Mismatches: 1
 Query Match: 87.42% Indels: 67
 DB: 20 Gaps: 3
 US-09-902-772-4 (1-478) x AAX26551 (1-1447)
 QY 1 MetalaserThrIleTyrGluAlaIleuSerValValSerGluAspGlnSerIleuPheGlu 20
 DB 63 ATGCCAAGCACTATTGAAGAAAGCATTTACAGTGTGAGTGAAGCAAGCTCTTTTGGAG 122
 QY 21 CysAlaIleTyrGlySerProHisLeuAlaIleTyrGluMetThrAlaSerSerSerGlu 40
 DB 123 TGTGCTCCACGAGATCGCCACCTTGCACAAAGACAGAAATGACAGCCTCTTCCAGTGAA 182
 QY 41 TyrGlyGlnThrSerLysMetSerProArgValProGlnGlnAspTyrPleuSerGlnPro 60
 DB 183 TATGGCAAAACATCAAAAGATGAGCCCGCGCTTCCCAAGCAGGACTGTATTCACAGCC 242
 QY 61 ProAlaArgValThrIleTyrMetGluCysAsnProAsnGlnValAsnGlySerArgAsn 80
 DB 243 CCGGCACAGATTACCATTAAGATGAGAGTGAACCAACGAGTTAATGGGTCAAGGAT 302
 QY 81 SerProAspAspCysSerValAlaIleGlyGlyLysMetValSerSerSerAspAsnVal 100

DB 303 TCACCTGATGACTCAGCGTGCCAAAGAGGAGAAATGTTAGCTTACAGACAACTT 362
 QY 101 GlyMetAsnTyrGlySerTyrMetGluGluLysHisIleProProAspMetThrThr 120
 DB 363 GGGATGAACATATGGAAGTACATGAGAGAGAGATATTCGCCCTCCCAATATATGCAAC 422
 QY 121 AsnGluArgValIleValProAlaAspProThrLeuTyrSerThrAspHisValArg 140
 DB 423 AATGACGAAGAGTTATGTCCTCCAGACATCTCTAGTTATGAGACACAGACCATATGACG 482
 QY 141 GlnThrPleuGluTyrPalaValIleGlyGluTyrGlyLeuProAspValAspIleLeuPhe 160
 DB 483 CAGTGGCTGAGAGTGGGAGTGAAGAGATGCTCTCCAGACGTGAGATCTGTGTTC 542
 QY 161 GlnAsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArgLeuThr 180
 DB 543 CAGAACATGATGGGAAAGAGTGTGTAAATGACCAAAAGATGACTTCAAGACTCAGC 602
 QY 181 ProSerTyrAsnAlaAspIleLeuLeuSerHisTyrLeuArgGluThrProLeu 200
 DB 603 CCGAGCTATACGCGAGATATCTCTGTCACACCTACACTACCTCAGAGAGA----- 656
 QY 201 ProHisLeuThrSerAspAspValAspLysAlaLeuGlnAsnSerProArgLeuMetHis 220
 DB 656 ----- 656
 QY 221 AlaArgAsnThrGlyAlaThrPheIlePheProAsnThrSerValTyrProGluAla 240
 DB 657 -----GAGGCCACTTTATTTTCCAAATATACATGATTACCCAGAACCA 701
 QY 241 ThrGlnArgIleThrThrArgProAspLeuProTyrGlnAlaIleArgSerAlaTyr 260
 DB 702 ACGCAAGAAATTAACACAGAGCCAGATTACCTTATGACAGCAGAGAGATCAGCG 761
 QY 261 ThrSerHisSerHisProThrGlnSerLysAlaThrGlnProSerSerThrValPro 280
 DB 762 ACGAGTCACACCATCCACACAGTCAAAAGCTCAACATCATCTTCAACAGTGC 821
 QY 281 LysThrGluAspGlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSerSer 300
 DB 822 AAAACAGAGCCAGCGCTCCTCAGTTAGATCTTATCATCTTCTTGACCGACCGACGC 881
 QY 301 ArgLeuAlaAsnProGlySerGlyIleGlnLeuTyrGlnPheLeuGluLeuLeu 320
 DB 882 CGTCTGCAAAATCCAGGAGTGGGAGATACAGTATGAGTCTCTGAGACTTCTG 941
 QY 321 SerAspSerSerAsnSerAsnCysIleThrTyrGluGlyThrAsnGlyGluPheLysMet 340
 DB 942 TCGACACAGCTCCAACTCCAACTGCATCACCTGGGAGGCGACAAATGGGAGTTCAAGATG 1001
 QY 341 ThrAspProAspGluValAlaArgArgTyrGluGlyLysSerLysProAsnMetAsn 360
 DB 1002 ACAGACCTGATGAAGTGGCTGGCGGTTGGGAGAGGAGAAAGCAAACTTACATGTGAC 1061
 QY 361 TyrAspLysLeuSerArgAlaLeuArgTyrTyrTyrAspLysAsnIleMetThrLysVal 380
 DB 1062 TATGCAAACTCAGCCGCGACCTTGGTACTACATATGCAAAATATATATACATAACTT 1121
 QY 380 ----- 380
 DB 1122 CATGTAACGCTATGCTTACAAATTTGATTTCCACGGAAATCGCTCAGCCCTCAGCCT 1181
 QY 381 HisProProGlnSerSerMetTyrLysTyrProSerAspLeuProTyrMetSerSerTyr 400
 DB 1182 CACCTCCAGATATCATCATATACAAATACCAATACAGACCTCCCTCATATAGTTCTTCA 1241
 QY 401 HisGlyLysArgTyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnPro 420
 DB 1241 ----- 1241
 QY 421 HisAlaHisProGlnLysMetAsnPheValAlaProHisProProAlaLeuProValThr 440

Db 1242 CAGCACACCCCAAGATGACTTTGAGCTCCCATCCCTGCTTGGCCGTAAAC 1301
 QY 441 SerSerSerPhePheAlaPheProAsnProTyrTyrPheSerProThrglylylyleTyr 460
 Db 1302 TCATCCAGCTTTTCTGCTGCTCAATCACTGATGATACCAACATGAGCCTACTAC 1361
 QY 461 ProAsnThrArgPheProAlaAlaIleMetProSerHisIleuGlyThrTyrTyr 478
 Db 1362 CCCAATACACAGCTGCGAGCTGCTCATATGCTTCCCATCTTGCGACCTACTAC 1415
 RESULT 3
 AA050644
 ID AA050644 standard: cDNA; 2938 BP.
 AC AA050644;
 XX
 DT 26-MAY-1994 (first entry)
 XX
 DE Human Hum-Flt-1 gene clone BM025.
 XX
 KW chromosomal translocation; chimeric; chimeric; Ewing sarcoma;
 KW Ews gene; malignant melanoma; hum-flt-1;
 KW primitive peripheral neuroectodermal tumour; human chromosome 11;
 KW human chromosome 22; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 143..1501
 FT /tag= a
 FT /product= HUM-FLT-1
 FT polyA_signal 2908..2913
 FT /tag= b
 FT
 FT
 XX WO9323549-A.
 PN 25-NOV-1993.
 PD
 XX 19-MAY-1993; 93WO-FR00494.
 PF
 XX 20-MAY-1992; 92PR-0006123.
 PR
 XX
 PA (CNRS) CNRS CENT NAT RECH SCI.
 XX
 PI Aurias A, Delattre O, Desmazaie C, Melot T, Peter M;
 PI Piongasel B, Thomas G, Zucman J;
 XX
 DR WPI; 1993-386580/48.
 DR P-PSDB; AAR44556.
 XX
 PT New nucleic acid of EWS gene and its hybrid(s) - contg. gene
 PT sequence involved in chromosomal translocation, also derived
 PT mRNA, probes, fusion proteins etc., for diagnosis and treatment
 PT of Ewing sarcoma and melanoma
 XX
 PS Disclosure; Fig 7; 123pp; French.
 XX
 CC The probe 11R1 was used to screen a human marrow cDNA library
 CC (Clontech cat. # HL1058). The clone BM025 was identified and
 CC sequenced. It represents the entire coding region together with
 CC 5'- and 3'-UTRs of the Hum-Flt-1 gene.
 XX
 SQ Sequence 2938 BP; 847 A; 692 C; 675 G; 724 T; 0 other;
 Alignment Scores:
 Pred. No.: 5.77e-113 Length: 2938
 Score: 1482.00 Matches: 291
 Percent Similarity: 68.65% Conservative: 55
 Best Local Similarity: 57.74% Mismatches: 80
 Query Match: 57.26% Indels: 78
 DB: 14 Gaps: 9
 US-09-902-772-4 (1-478) x AA050644 (1-2938)

QY 1 MetAlaSerThrIleuylSerGluAlaLeuSerValIleSerGluAspIleuSerLeuPheGlu 20
 Db 143 ATGAGGGGAGCTATTAAGGAGGCTGCTGCTGTGAGGACACACAGCCCTTGTGAC 202
 QY 21 CysAlaTyrGly---SerProHisIleuAlaIleTyrThrGluMetThrAlaSerSerSer 39
 Db 203 TCAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 252
 QY 40 GluTyrGlyGlnThrSerIleuSerMetSerProArgValIleProGlnIleuSerPheSerGln 59
 Db 263 GACTCTCCGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG 322
 QY 60 ProProAlaArgValThrIleuSerMetGluCysAsnProAsnGlnValAsnGlySerArg 79
 Db 323 ---CCAGTGGAGGTCAACGTCAGCGAGGTAT---GACCACATGATGATGATCAG 373
 QY 80 AsnSerProAspAspCysSerValAlaIleValGlyIleMetValSerSerSerAspAsn 99
 Db 374 GAGCTCCGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG 433
 QY 100 ValGlyMetAsnTyrGlySerTyrMetGluGluIleHis---IleProProAsnMet 118
 Db 434 AACCCCATGACTACCAAGCTATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 493
 QY 119 ThrThrAsnGluArgArgValIleValProAlaAspProThrLeuThrSerThrAspHis 138
 Db 494 ACCACCAAGAT 553
 QY 139 ValArgGlnThrPheGlnIleThrPheValValGlyIleuProAspValIleu 158
 Db 554 GTAGGCAATGCTGAGAGTGGCCATTAAGAGAGATATAGCTTGTAGAGATCGACACATCC 613
 QY 159 LeuPheGlnAsnIleAspGlyIleuGluLeuGlySerMetThrIleuAspAspPheGlnArg 178
 Db 614 TTTTCCAGAACATGATGAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 673
 QY 179 LeuThrProSerTyrAsnAlaAspIleuLeuSerHisIleuHisTyrLeuArgGluThr 198
 Db 674 GCCACCACCTCTCAACAG 733
 QY 199 ProLeuProHisIleuThrSerAspValAspValAlaLeuGlnAsnSerProArgLeu 218
 Db 734 TCAGCTG-----CTGGCTATATCAACCTCCACAC 739
 QY 219 MetHisAlaArgAsnThrGlyAlaThrPheIlePheProAsnThrSerValTyrPro 238
 Db 740 -----CTGGCTATATCAACCTCCACAC 766
 QY 239 GluAlaThrGlnArgIleThrThrArgProAspLeuProTyrGlnGluAlaArgArgSer 258
 Db 767 GACCAATCTCCACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 826
 QY 259 AlaThrThrSerHisSerHisProThrGlnSerValAlaThrGln---ProSerSerSer 277
 Db 827 GCATGGGGCAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 886
 QY 278 ThrValProIleuThrGluAspGlnArgProGlnIleuAspProTyrGlnIleuGlyPro 297
 Db 887 ACATATAGTAAGAAATACAG 946
 QY 298 ThrSerThrArgLeuAlaAsnProGlySerGlyGlnIleGlnLeuThrPheLeuLeu 317
 Db 947 ACCAGAGAGTCCCTACCAACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1006
 QY 318 GluIleuLeuSerAspSerSerAsnSerAsnValIleThrTyrGlnIleuThrAsnGlyGlu 337
 Db 1007 GAGCTGCTCTCCACAG 1066
 QY 338 PheIleuMetThrAspProAspGluValAlaArgArgTyrGlyIleuArgIleuSerIleuPro 357
 Db 1067 TTCAAATATGAG 1126

QY 358 AsnMetAsnTrpAspLysLeuSerArgAlaLeuArgTyrTyrTrpAspLysAsnIleMet 377
 Db 1127 AACGTGATTAACAGCAAGCTGAGCGGCCCTCCGTTATTACTATGATTAATAAATATG 1186
 QY 378 ThrTyrValHisProProGluSerSerMetTyrTyrProSerAspLeuProTyrMet 397
 Db 1187 ACCAAATG----- 1195
 QY 398 SerSerTyrHisGlyLysArgTyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAla 417
 Db 1196 -----CACGGCAAAAGATATGTTACAAATTTGACTCCAGCGCATGGCCAGGCT 1246
 QY 418 LeuGlnPro----- 420
 Db 1247 CTGAGGCAATCCGAGCGAGTCTCATGTACAGTACCCTTCTGACATCTCCATCATG 1306
 QY 421 -----HisAlaHisProGlnLysMetAsnPheValAlaProHisProProAlaLeu 437
 Db 1307 CCTTCCATCATGCGCCACGACAGAGTGAATCTTGTCCCTCCATCCATCTCCATG 1366
 QY 438 ProValThrSerSerSerPhePheAlaAlaProAsnProTyrTyrPheAsnSerProThrGly 457
 Db 1367 CCTGTCACTTCCTCAGCTCTTGTGGAGCGCATCACAAATAGTGGACCTCCGCCAGGGG 1426
 QY 458 GlyIleTyrProAsn-----ThrArgLeuProAlaAlaHisMetProSerHisLeu 474
 Db 1427 GGAATTCACCCCAACCCCAAGTCCCGCCATCTTAACACCCAGCGCTTCACACTTA 1486
 QY 475 GlyThrTyrTyr 478
 Db 1487 GCGAGCTACTAC 1498
 RESULT 4
 ID ABR84139 standard; cDNA; 2957 BP.
 XX ABR84139;
 AC ABR84139;
 DT 14-AUG-2002 (first entry)
 DE Human cDNA differentially expressed in granulocytic cells #710.
 XX
 KW Human; ss; granulocytic cell; DNA chip; bacterial infection;
 KW viral infection; parasitic infection; protozoal infection;
 KW fungal infection; sterile inflammatory disease; psoriasis;
 KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
 KW cardiac reperfusion injury; renal reperfusion injury; ARDS;
 KW adult respiratory distress syndrome; inflammatory bowel disease;
 KW Crohn's disease; ulcerative colitis; periodontal disease;
 KW granulocyte activation; chronic inflammation; allergy.
 XX
 OS Homo sapiens.
 PN W0200228999-A2.
 PD 11-APR-2002.
 XX
 PF 03-OCT-2001; 2001WO-US30821.
 XX
 PR 03-OCT-2000; 2000US-237189P.
 XX
 PA (GENE-) GENE LOGIC INC.
 XX
 PI Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;
 XX
 DR WPI; 2002-435328/46.
 XX
 PT Detecting granulocyte activation by detecting differential expression
 of genes associated with granulocyte activation, which serves as
 PT diagnostic markers that is useful for monitoring disease states and
 drug toxicity -
 XX
 PS Claim 1; SEQ ID NO 710; 114bp; English.

XX The invention relates to detecting (M1) granulocyte (GC) activation
 CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by
 CC DNA chip analysis as given in the specification, and comparing
 CC the expression level to an expression level in an unactivated
 CC GC, where differential expression of Gs is indicative of GCA.
 CC Also included are modulating (M2) GA by contacting GC with an agent
 CC that alters the expression of at least one gene in Gs; (2) screening (M3)
 CC for an agent capable of modulating GCA or an inflammatory (especially
 CC chronic) in a tissue, an allergic response in a subject, exposure of a
 CC subject to a pathogen or sterile inflammatory disease using the
 CC gene expression profile; (3) detecting (M4) an inflammation (especially
 CC chronic) in a tissue, an allergic response in a subject, exposure of a
 CC subject to a pathogen or sterile inflammatory disease, by detecting the
 CC level of expression in a sample of the tissue of gene(s) from Gs, where
 CC (4) level of expression of the gene is indicative of inflammation;
 CC an allergic response in a subject, exposure of a subject to a pathogen
 CC or sterile inflammatory disease, by contacting a tissue having
 CC inflammation with an agent that modulates the expression of gene(s)
 CC from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for
 CC modulating GA; M3 is useful for screening an agent capable of modulating
 CC GCA preferably in an inflammation in a tissue; M4 is useful for
 CC detecting an inflammation (especially chronic) in a tissue, an allergic
 CC response in a subject, exposure of a subject to a pathogen or sterile
 CC inflammatory disease (e.g. psoriasis, rheumatoid arthritis,
 CC glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal
 CC reperfusion injury, ARDS, adult respiratory distress syndrome,
 CC inflammatory bowel disease, Crohn's disease, ulcerative colitis,
 CC periodontal disease; also bacterial infection, viral infection,
 CC parasitic infection, protozoal infection, fungal infection and M5 is
 CC useful for treating one of the above conditions. The present
 CC sequence represents a gene differentially expressed in granulocytes.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 2957 BP; 855 A; 698 C; 680 G; 724 T; 0 other;
 Alignment Scores:
 Pred. No.: 5.82e-113 Length: 2957
 Score: 1482.00 Matches: 291
 Percent Similarity: 68.65% Conservative: 55
 Best Local Similarity: 57.74% Mismatches: 80
 Query Match: 57.26% Indels: 78
 DB: 24 Gaps: 9
 US-09-902-772-4 (1-478) x ABR84139 (1-2957)
 QY 1 MetaLaserThrIleLysGluAlaLeuSerValAlaSerGluAspGlnSerLeuPheGlu 20
 Db 173 ATGACGCGGACTATTAAAGAGGCTGTGCTGGTGACGACGACCACTCCCTTTGAC 232
 QY 21 CysAlaTyrGly--SerProHisLeuAlaLysThrGlnMetThrAlaSerSerSer 39
 Db 233 TCACGCTAGCGAGGCGGAGCCCATCTCCCAAGGCCGACATGACTGCTCGGGAGTCTT 292
 QY 40 GlnTyrGlnGlnThrSerLysMetSerProArgValProGlnGlnAspTrpLeuSerGln 59
 Db 293 GACTACGGCGACCCACAGATCAACCCCTCCACACAGCAGAGTGATCAATCAG 352
 QY 60 ProProAlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArg 79
 Db 353 ---CCAGTGAAGGCTCAACGTCACAGCGGAGAT-----GACCAATGATGATCCAGG 403
 QY 80 AsnSerProAspAspCysSerValAlaLysGlyLysMetValSerSerSerAspAsn 99
 Db 404 GAGCTCTCGGTGACTCAGCGGTGACAAATGACAGACGTGGGCGGAGGAGTCC 463
 QY 100 ValGlyMetAsnTyrGlySerTyrMetGlnLysHis---IleProProProAsnMet 118
 Db 464 AACCCATGAACATCAACAGCTATATGAGCAGAGAAGATGCCGCCCTCTCCCAACATG 523

```

QY 119 ThrThrAsnGlnArgArgValIleValProAlaAspProThrLeuThrSerThrAspHis 138
    |||||
DB 524 ACACCAACAGAGAGAGATCATCTCCCGCAGACCCACACTGTGGACACAGAGCAT 583
QY 139 ValArgGlnThrLeuGlnThrPheAlaValLysGlnTyrGlyLeuProAspValAspIleLeu 158
    |||||
DB 584 GTGAGGCAATGGCTGGAGGCGCAATGAAGAGTACAGCTGTGTGAGATCGACACATCC 643
QY 159 LeuPheGlnAsnIleAspGlyLysGlnLeuCysLysMetThrLysAspAspPheGlnArg 178
    |||||
DB 644 TTTTCCAAACATGATGATGCAAGAGAACTGTGTAAATGATACAGAGAGACTCTCCCGC 703
QY 179 LeuThrProSerThrArgAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGlnThr 198
    |||||
DB 704 GCCACCACTCTACACACGAGAGTGTGTGTCTACACTCTCAGTTACCTCAGGAAAGT 763
QY 199 ProLeuProHisLeuThrSerAspAspValAspLysAlaLeuGlnAsnSerProAlaGlyLeu 218
    |||
DB 764 TCACGTG----- 769
QY 219 MethIAsIaArgAsnThrGlyAlaIleThrPheIlePheProAsnThrSerValTyrPro 238
    |||||
DB 770 -----CTGGCCTATATACAACTCCACACC 796
QY 239 GluAlaThrGlnArgIleThrThrArgProAspLeuProTyrGlnAlaAlaArgArgSer 258
    |||||
DB 797 GACCAATCCTCAGATGTGTCAAGAAGAACCTTCTATGACACAGTCAAGAAAGAGA 856
QY 259 AlaThrThrSerHisSerHisProThrGlnSerHisAlaThrGln---ProSerSerSer 277
    |||||
DB 857 GCTTGGGCAATACATACATGATCTGGCCCTCAACAAATGCTCCCTTGAGAGGGGCAAA 916
QY 278 ThrValProLysThrLysGlnArgGlnArgProGlnLeuAspProTyrGlnIleLeuGlyPro 297
    |||||
DB 917 ACATAGATGAAGAATACAGACAGACAGCCCGCCAGATCCGATAGATCCGAGGCCG 976
QY 298 ThrSerSerArgLeuAlaAsnProGlySerGlyGlnIleGlnLeuThrPheLeuLeu 317
    |||||
DB 977 ACCAGAGTCGCTTACCAACCTCGAAGCGGGCAGATCAGCTGTGCAATTCCTCTCG 1036
QY 318 GluLeuLeuSerSerSerSerAsnSerAsnGlyIleThrThrGlnGlnThrAsnGlyGln 337
    |||||
DB 1037 GAGCTGCTCTCCAGACGCCCAACGCGACGTGATCACTGGAGGGGAGCAACGGGGAG 1096
QY 338 PheLysMetThrAspProAspGlnValAlaArgArgTyrGlyGlnArgLysSerLysPro 357
    |||||
DB 1097 TTCAAAATGACGGACCCGATGAGTGGCCAGGCGCTGGGGGAGCGGAAAGCAAGCC 1156
QY 358 AsnMetAsnTyrAspLysLeuSerArgAlaLeuArgTyrTyrTyrAspLysAsnIleMet 377
    |||||
DB 1157 AACATGAATTAGACAAAGCTGAGCCCGGCCCTCCGTTATCTATCATATAAACAATTAG 1216
QY 378 ThrLysValHisProProGlnSerMetTyrLysTyrProSerAspLeuProTyrMet 397
    |||||
DB 1217 ACCAAAGTC----- 1225
QY 398 SerSerTyrHisGlyLysArgTyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAla 417
    |||||
DB 1226 -----CACGGCAAAAGATATGCTTACAAATTGATTCACGGCATTTGCCAGGCT 1276
QY 418 LeuGlnPro----- 420
    |||||
DB 1277 CTGCACCCACATCCGACGAGTCGTCATGACAACTACCCCTTGACATCTCTACATG 1336
QY 421 -----HisAlaHisProGlnLysMetAsnPheValAlaProHisProProAlaLeu 437
    |||||
DB 1337 CCTTCTACCATGCTCCACAGAGAGGAGAACTTTGCTCTCCCTCCATCCATCCATCCAG 1396
QY 438 ProValThrSerSerSerPhePheAlaAlaProAsnProTyrTyrPheAsnSerProThrGly 457
    |||||
DB 1397 CCGTGTACTTCTCCAGCTTCTTTTGGAGCGGCATCAATACTGGAGCTCCGCCACGGGG 1456

```

```

QY 458 GlyIleTyrProAsn-----ThrArgLeuProAlaAlaHisMetProSerHisLeu 474
    |||||
DB 1457 GGAAATCTACCCCAACCAACGTCCTCCCGCATCTTACACCCACGCTGCTTACACTTAA 1516
QY 475 GlyThrTyrTyr 478
    |||||
DB 1517 GGCAGCTACTACT 1528

RESULT 5
AAH02915
ID AAH02915 standard; DNA; 2954 BP.
AC AAH02915;
DC 15-JUN-2001 (first entry)
DE Human shear stress-response coding sequence SEQ ID NO: 83.
KW Human; shear stress-response protein; vascular disease;
KW arteriosclerosis; ds.
XX
XX Homo sapiens.
OS
PN MO200125427-A1.
XX
XX 12-APR-2001.
PD
XX
XX 02-OCT-2000; 2000WO-JP06840.
PF
XX
XX 01-OCT-1999; 99JP-0280976.
PA (KIOM ) KIOMA HAKKO KOYO KK.
PA (NOJI/) NOJIMA H.
PI Nojima H, Yoshisue H, Obayashi M, Ota T, Kawabata A, Sakurada K,
PI Kuga T, Sekine S, Nakamura Y, Sugano S;
DR WPI: 2001-266308/27.
DR P-PSDB: AAB90792.
XX
XX DNA sequences, proteins encoded by them and antibodies against them
PT useful in diagnosis and treatment of vascular disease caused by
PT arteriosclerosis -
XX
XX PS Claim 20; Page 462-466; 678bp; Japanese.
XX
XX CC The present invention provides the protein and coding sequences of a
CC number of human shear stress response proteins. These are useful in the
CC diagnosis, treatment and screening of vascular diseases caused by
CC arteriosclerosis, including heart failure, post-PTCA restenosis and
CC hypertension.
XX
SQ Sequence 2954 BP; 846 A; 698 C; 683 G; 727 T; 0 other;

Alignment Scores:
Pred. No.: 1,64e-111 Length: 2954
Score: 1464.50 Matches: 290
Percent Similarity: 68.45% Conservative: 55
Best Local Similarity: 57.54% Mismatches: 80
Query Match: 56.59% Indels: 79
DB: 22 Gaps: 10

US-09-902-772-4 (1-478) x AAH02915 (1-2954)
QY 1 MetAlaSerThrIleLysGlnAlaLeuSerValValSerGluAspGlnSerLeuPheGln 20
    |||
DB 173 ATGAGCGGAGCTATTAGAGGCTGTGTGGTGTAGCGACGACGATCCCTCTTTGAC 232
QY 21 CysAlaTyrGly---SerProHisLeuAlaLysThrGluMetThrAlaSerSerSer 39
    |||||
DB 233 TCAGCTTACGAGAGCGGACCATCTCCCAAGCGCCGACATGATGCTCGGGGATCCT 292
QY 40 GluTyrGlyGlnThrSerLysMetSerProArgValProGlnGlnAspTyrLeuSerGln 59

```

```

Db 293 GACATCGGGGACCCCAACATGATCAACCCCTCCACACAGCAGAGGTGATCAATCAG 352
Qy 60 ProProAlaArgValThrIleLeuMetGluCysAnPProAsnGlnValAsnGlySerArg 79
Db 353 ---CCAGTGAAGGCTCAACGTCAACGGGAGAT-----GACCAACATGAATGATCCACG 403
Qy 80 AsnSerProAspAspCysSerValAlaLysGlyLysMetValSerSerSerAspAsn 99
Db 404 GAGTCTCCGCTGAGCTGACGTGACGATAAGCAAAATGACAGCAAGTGGTGGCGAGGAGATCC 463
Qy 100 ValGlyMetAsnTyrGlySerTyrMetGlnLysHis-----IleProProProAsnMet 118
Db 464 AACCCCAATGAACCTACACAGCTATATGACGAGAAGATGGCCCCCTCCCTCCCAACATG 523
Qy 119 ThrThrAsnGlnArgArgValIleValProAlaAspProThrLeuTyrSerThrAspHis 138
Db 524 ACCACCAACGAGAGAGAGATGATGTCGCCACACCCACACGTGTGACACAGAGAGAT 583
Qy 139 ValArgGlnTyrLeuGluTyrPalaValLysGluTyrGlyLeuProAspValAspIleLeu 158
Db 584 GTGAGGCAATGGCTGAGTGGCCATTAAGAGATACAGCTGATGAGATGCACATCC 643
Qy 159 LeuPheGlnAsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArg 178
Db 644 TTTTTCAGAACATGATGAGCAGAGCAACTGTGTAAAGAAACAAGAGAGACTTCCTCCGC 703
Qy 179 LeuThrProSerTyrAsnAlaAspIleLeuLeuSerHisLysLysTyrLeuArgGluThr 198
Db 704 GCCACCAACCCCTACACACGAGAGTGTGTGTGCACACCTCACTTACCTCAGGAAAGT 763
Qy 199 ProLeuProHisLeuThrSerAspValAspLysAlaLeuGlnAsnSerProArgLeu 218
Db 764 TCACGTG----- 769
Qy 219 MethIslaArgAsnThrGlyAlaThrPheIlePheProAsnThrSerValTyrPro 238
Db 770 -----CTGGCCTATATATACAACTCCACACC 796
Qy 239 GluAlaThrGlnArgIleThrThrArgProAspLeuProTyrGlnAlaIlaArgSer 258
Db 797 GACCAATCCCTACAGATGAGTCAAGAAGAACCCCTTATGACTCAGTCAGAGAAGA 856
Qy 259 AlaThrPheSerHisSerHisProThrGlnSerLysAlaThrGln---ProSerSerSer 277
Db 857 GCTTGGGGCAATACATGAATCTGGCCCTCAACAAAGTCCCTCCCTTGAGGGGACAA 916
Qy 278 ThrValProLysThrGluAspGlnArgProGlnLeuAspProTyrGlnIleLeuGlyPro 297
Db 917 ACGATCAGTAAAGATACAGAGCAAGCGCCAGCAATCCGTATCAGATCCTGGCCCG 976
Qy 298 ThrSerSerArgLeuAlaAsnProGlySerGlyGlnIleGlnLeuTyrPglPheLeuLeu 317
Db 977 ACCAGCAGTCCCTACGCAACCTCGGAAGCGGAGATCCAGCTGTGCAATTCCTCTCG 1036
Qy 318 GluLeuLeuSerAspSerSerAsnSerAsnGlyIleThrTyrGluGlyThrAsnGlyGlu 337
Db 1037 GAGCTGCTCTCCGACAGGCGCACGCCAGCTGATCACTGAGGGGAGACCAAGGGAG 1096
Qy 338 PheLysMetThrAspProAspGluValAlaArgArgTyrPglGlyLysGlySerLysPro 357
Db 1097 TTCAAATGACGAGCCCGGATGAGTGGCCGCGGCTGGGGCGGCGAAAGCAAGCC 1156
Qy 358 AsnMetAsnTyrAspLysLysLeuSerArgAlaLeuArgTyrTyrAspLysAsnIleMet 377
Db 1157 AACATGAATATGACAAAGCTACGCCGGCCCTCCGTTATATCTATGATATAAAACATTATG 1216
Qy 378 ThrLysValHisProProGluSerSerMetLysTyrProSerAspLeuProTyrMet 397
Db 1217 ACCAAAGTG----- 1225
Qy 398 SerSerTyrHisGlyLysArgTyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAla 417

```

```

Db 1226 -----CACGCGAAAGATATGCTTACAAATTTGACTTCCACAGGCATTTGCCAGGCT 1276
Qy 418 LeuGlnPro----- 420
Db 1277 CTCGACGACCATCCGACGAGTGCATGTACAGTACCTCTGTGACATCTCTACATG 1336
Qy 421 -----HisAlaHisProGlnLysMetAsnPheValAlaProHisProProAlaLeu 437
Db 1337 CCTTCCCAACATGGCCACACAGAGGTGAATTTGTCCCTCCCATCATCTCCATG 1396
Qy 438 ProValThrSerSerPhePheAlaIlaProAsnProTyrTyrPAsnSerProThrGly 457
Db 1397 CCTGTCACTTCTCCAGCTTCTTGGAGCCGATCACAATCTGGACCTC---ACGGGG 1453
Qy 458 GlyIleTyrProAsn-----ThrArgLeuProAlaIlaHisMetProSerHisLeu 474
Db 1454 GGAAATCTACCCCAACCCCAAGCTCCCGCATCTCTATACACGACGTCCTTACACTTA 1513
Qy 475 GlyThrTyrTyr 478
Db 1514 GGCAGCTACTAC 1525

RESULT 6
AA050662 standard; DNA; 1890 BP.
AC
XX
AA050662;
XX
DE 26-MAY-1994 (first entry)
XX
XX Human Hum-Fil-1 gene (genomic DNA).
XX
KW Chromosomal translocation; chimeric; chimaeic; Ewing sarcoma;
KW Ews gene; malignant melanoma; hum-fil-1;
KW primitive peripheral neuroectodermal tumour; human chromosome 11;
KW human chromosome 22; ss.
XX
XX Homo sapiens.
XX
OS
XX
FH Key 1.39 location/Qualifiers
FT exon
FT FT //tag= a
FT //number= 1
FT //codon_start= 22..24
FT FT //tag= b
FT //number= 1
FT FT //number= 100..311
FT exon
FT FT //tag= c
FT //number= 2
FT FT //tag= d
FT //number= 2
FT FT //number= 372..526
FT exon
FT FT //tag= e
FT //number= 3
FT FT //tag= f
FT //number= 527..586
FT FT //tag= g
FT //number= 4
FT FT //tag= h
FT //number= 4
FT FT //tag= i
FT //number= 851..916
FT exon
FT FT //tag= j
FT //number= 5
FT FT //tag= k
FT //number= 977..1042
FT FT //tag= k

```



```
QY 294 ILeuGlyProThrSerSerArgLeuAlaPro----- 305
Db 1234 ATCTGGGCGGACGACAGTGCCTAGCAACCTGGTGAAGTTACCTGGCTGCAAG 1293
QY 306 -----GlySerGlyGlnIleGlnLeuTrp 313
Db 1294 CCTTTTGTCTCTCCCGTTTCTCAAGCGGCTGACAGAAAGCGGAGATCCAGCTGTGG 1353
QY 314 GlnPheLeuLeuGlnLeuLeuSerAspSerSerAsnSerAsnCysIleThrTrpGluGly 333
Db 1354 CAATTCCTCGTGGAGCTGCTCTCCGACAGGCGCAAGCCAGCTATACACTGGAGAGGG 1413
QY 334 ThrAsnGlyGlnPheLeuMetThrAspProAspGluValAlaIleGlnGlyGluArg 353
Db 1414 ACCAACGGGGGAGTTCAAAATGACGACCCGAGTGAAGTGGCGGCGCTGGGGGAGCGG 1473
QY 354 LysSerLysProAsnMetAsnTrpAspLysLeuSerArgAlaLeuAlaGlyTrpTyrAsp 373
Db 1474 AAAAGCAAGCCCAACATGATGATTACGACAGCTGAGCGGCGCTTATTAATGAT 1533
QY 374 LysAsnIleMetThrLysValHisProGluSerSerMetLysTrpTyrProSerAsp 393
Db 1534 AAAAACATTAATGACCAAGTG----- 1554
QY 394 LeuProGlyMetSerSerTrpHisGlyLysArgTyrAlaTyrLysPheAspPheHisGly 413
Db 1555 -----CACGGCAAAAGATATGCTTACAAATTTGACTTCACGCGC 1593
QY 414 IleAlaGlnAlaLeuGlnPro----- 420
Db 1594 ATTGCCAGGCTCTGACGCCACATCCGACGAGTCGTCATGTCAGATCCCTTTCAC 1653
QY 421 -----HisAlaHisProGlnLysMetAsnPheValAlaProHis 433
Db 1654 ATCTCTACATGCTTCTTACATGACGACGACGAGAGGAGTTCCTGCCCTCCCAT 1713
QY 434 ProProAlaLeuProValThrSerSerSerPhePheAlaProAsnProTyrTrpAsn 453
Db 1714 CCATCCCTCAGCTGCTGATCTCTCCAGCTTCTTGAGCGCGATCAATATCTGAGCC 1773
QY 454 SerProThrGlyGlyIleTyrProAsn-----ThrArgLeuProAlaAlaHisMet 470
Db 1774 TCCCCCAGGGGGGAAATCTACCCCAACCCAGACGTCCCGCATCTTACACCCAGGTG 1833
QY 471 ProSerHisLeuGlyThrTyr 478
Db 1834 CCTTCACACTTAGGACACTACTAC 1857

RESULT 7
ABA48124/c
ID ABA48124 standard; DNA; 567 BP.
XX
XX ABA48124:
XX
XX 01-FEB-2002 (first entry)
XX
XX Human breast cell single exon nucleic acid probe #6819.
XX
XX Human: microarray; single exon probe; gene expression; breast;
XX
XX disease; cancer; ss.
XX
XX Homo sapiens.
XX
XX MO200157271-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001MO-US00662.
XX
XX 04-FEB-2000; 2000US-0180312.
XX
XX 26-MAY-2000; 2000US-0207456.
XX
XX 30-JUN-2000; 2000US-0608408.
XX
XX 03-AUG-2000; 2000US-0632366.
```

```
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-496933/54.
XX
XX New spatially-addressable set of single exon nucleic acid probes,
XX
XX useful for measuring gene expression in sample derived from human
XX
XX breast, comprises number of single exon nucleic acid probes
XX
XX Claim 4; SEQ ID NO 6819; 327pp + sequence listing; English.
XX
XX The invention relates to a spatially-addressable set of single exon
XX
XX nucleic acid probes for measuring gene expression in a sample derived
XX
XX from human breast and BT 474 cells. The method involves contacting
XX
XX the probes with a collection of detectably labelled nucleic acids
XX
XX derived from mRNA of human breast, and then measuring the label
XX
XX bound to each probe of the microarray. The probes are useful for
XX
XX verifying the expression of regions of genomic DNA predicted to
XX
XX encode proteins. They are useful for gene discovery, and for
XX
XX determining predisposition and/or prognosing breast disease. Gene
XX
XX expression analysis is useful for assessing the toxicity of chemical
XX
XX agents on cells. The microarray of this invention presents a far greater
XX
XX diversity of probes for measuring gene expression, with far less bias
XX
XX than expressed sequence tag microarrays. The method is suitable for
XX
XX rapid production of functional information from genomic sequence. The
XX
XX present sequence is a single exon nucleic acid probe of the invention.
XX
XX Note: The sequence data for this patent did not form part of the
XX
XX printed specification, but was obtained in electronic format directly
XX
XX from WIPO at ftp.wipo.int/pub/published_pcl_sequences.
XX
XX SQ Sequence 567 BP; 114 A; 124 C; 200 G; 129 T; 0 other:
XX
XX Alignment Scores:
XX
XX Pred. No.: 1,78e-57 Length: 567
XX
XX Score: 801.00 Matches: 151
XX
XX Percent Similarity: 78.17% Conservative: 3
XX
XX Best Local Similarity: 76.65% Mismatches: 3
XX
XX Query Match: 30.95% Indels: 40
XX
XX DB: 22 Gaps: 2
XX
XX US-09-902-772-4 (1-478) x ABA48124 (1-567)
XX
XX QY 302 LeuAlaAsnProGlySerGlyGlnIleGlnLeuTrpGlnPheLeuGlnLeuSer 321
XX
XX Db 534 CTGACCTCATTTAGGACAGTGGCCAGATCCAGCTTGGCAGTTCCTCGAGACTCTGCG 475
XX
XX QY 322 AspSerSerAsnSerAsnCysIleThrTrpGlnGlyThrAsnGlyGlnPheLeuMetThr 341
XX
XX Db 474 GACAGCTCCAGCTCCAGCTGATCATCCTGGGAAGCACCAAGGAGTTAAAGATGACG 415
XX
XX QY 342 AspProAspGluValAlaIleArgTrpGlyGlnArgLysSerLysProAsnMetAsnTrp 361
XX
XX Db 414 GATCCCGACGAGTGGCGCGCTGGGGAGAGGGAAGACCAACCAATCAATCACTAC 355
XX
XX QY 362 AspLysLeuSerArgAlaLeuArgTyrTyrAspLysAsnIleMetThrLysValHis 381
XX
XX Db 354 GATAGCTACAGCCCGCGCTCCGTTACTACTATACAAAGAACATCATGACCAAGTGC-- 298
XX
XX QY 382 ProProGluSerSerMetLysTrpTyrProSerAspLeuProTyrMetSerSerTrpHis 401
XX
XX Db 297 -----CAT 295
XX
XX QY 402 GlyLysArgTyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnPro--- 420
XX
XX Db 294 GGAAGGCGTACGCTTCAAGTTGAGCTTCCAGGGAGTCCGCGCTTCCAGCCCCAC 235
XX
XX QY 421 -----His 421
```

```

Db      234 CCCCCGAGTCATCTGTGACAAAGTACCCCTCAGACCTCCCGTACATGGGCTCTATTCAC 175
Oy      422 AAlaHisProGlnLysMetAsnPheValAlaProHisProProAlaLeuProValThrSer 441
      174 GCCCACCACAGAGATGAACTTTGTGGGCGCCACCTCCAGCCCTCCCGCGATATCT 115
Oy      442 SerSerPhePheAlaAlaProAsnProTyrTTPasSerProThrGlyIleTyrPro 461
      114 TCCAGTTTTTTTGTCTCCCAACCCATCTGGAATTCACCACTGGGGGTATATATACCC 55
Db      462 AsnThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr 478
      54 AACACTAGGCTCCCCACGACCATATGCTTCTCATCTGTGGGCGCTACTATAC 4
RESULT 8
ABa66003/c
ID      ABa66003 standard; DNA; 567 BP.
XX
XX      AC      ABa66003;
XX      DT      01-FEB-2002 (first entry)
XX      DE      Human foetal liver single exon nucleic acid probe #14308.
XX      KW      Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX      OS      Homo sapiens.
XX      PN      WO200157277-A2.
XX      PD      09-AUG-2001.
XX      PF      30-JAN-2001; 2001WO-US00669.
XX      PR      04-FEB-2000; 2000US-0180312.
XX      PR      26-MAY-2000; 2000US-0207456.
XX      PR      30-JUN-2000; 2000US-0608408.
XX      PR      03-AUG-2000; 2000US-0632366.
XX      PR      21-SEP-2000; 2000US-0234687.
XX      PR      27-SEP-2000; 2000US-0236359.
XX      PR      04-OCT-2000; 2000GB-0024263.
XX
XX      (MOLE-) MOLECULAR DYNAMICS INC.
XX      PI      Penn SG, Hanzel DK, Chen W, Rank DR;
XX      DR      WPI; 2001-483447/52.
XX      PT      Human genome-derived single exon nucleic acid probes useful for
XX      PT      analyzing gene expression in human foetal liver.
XX      PS      Claim 4; SEQ ID NO 14308; 639pp + sequence listing; English.
XX      CC      The invention relates to a single exon nucleic acid probe for
XX      CC      measuring human gene expression in a sample derived from human foetal
XX      CC      liver. The single exon nucleic acid probes may be used for predicting,
XX      CC      measuring and displaying gene expression in samples derived from human
XX      CC      foetal liver. The present sequence is a single exon nucleic acid
XX      CC      probe of the invention.
XX      CC      Note: The sequence data for this patent did not form part of the
XX      CC      printed specification, but was obtained in electronic format directly
XX      CC      from WIPO at ftp.wipo.int/pub/published_pat_sequences.
XX      SQ      Sequence 567 BP; 114 A; 124 C; 200 G; 129 T; 0 other;
Alignment Scores:
Pred. No.:      1.78e-57      Length:      567
Score:          801.00      Matches:      151
Percent Similarity: 78.17%      Conservative: 3
Best Local Similarity: 76.65%      Mismatches: 3
Query Match:    30.95%      Indels:      40
DB:            22      Gaps:        2

```

```

US-09-902-772-4 (1-478) x ABa66003 (1-567)
Oy      302 LeuAlaAsnProGlySerGlyGlnIleGlnLeuThrPheLeuLeuGlnLeuLeuSer 321
      534 CTGACCTCATTAAGCAGTGCGCCAGATCCAGCTTTGGCAGTCTCTCTGGAGGCTCGTGCG 475
Oy      322 AspSerSerAsnSerAsnGlyIleThrTPGluGlyThrAsnGlyIleGlyMetThr 341
      474 GACAGCTCCACATCCCGACGATCACCCTGGGAAGGACCAACGGGAGTTCAGATGACAG 415
Oy      342 AspProAspGluValAlaAlaArgArgTyrGlyIleArgGlySerLysProAsnMetAsnTyr 361
      414 GATCCGACGAGAGGTGGCCCGCGCTGGGAGACCGGAACCAACCAACATGAAATCAC 355
Oy      362 AspLysLeuSerArgAlaLeuArgTyrTyrTyrAspLysAsnIleMetThrLysValHis 381
      354 GATTAAGCTCAGCGCGCCCTCGCTTACTACTATGACAGACATCATGACCAAGATC--- 298
Oy      382 ProProGluSerSerMetTyrTyrTyrProSerAspLeuProTyrMetSerTyrHis 401
      297 -----CAT 295
Oy      402 GlyLysArgTyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnPro--- 420
      294 GGGAAAGCCTACGCTTACAAAGTTTCGACTTCACGGGATGCCAGGCCCTCCAGCCCCAC 235
Oy      421 -----His 421
Db      234 CCCCCGAGTCATCTGTGACAAAGTACCCCTCAGACCTCCCGTACATGGGCTCTATTCAC 175
Oy      422 AAlaHisProGlnLysMetAsnPheValAlaProHisProProAlaLeuProValThrSer 441
      174 GCCCACCACAGAGATGAACTTTGTGGGCGCCACCTCCAGCCCTCCCGCGATATCT 115
Oy      442 SerSerPhePheAlaAlaProAsnProTyrTTPasSerProThrGlyIleTyrPro 461
      114 TCCAGTTTTTTTGTCTCCCAACCCATCTGGAATTCACCACTGGGGGTATATATACCC 55
Oy      462 AsnThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr 478
      54 AACACTAGGCTCCCCACGACCATATGCTTCTCATCTGTGGGCGCTACTATAC 4
RESULT 9
ABa33090/c
ID      ABa33090 standard; DNA; 567 BP.
XX
XX      AC      ABa33090;
XX      DT      23-JAN-2002 (first entry)
XX      DE      Probe #11556 for gene expression analysis in human heart cell sample.
XX      KW      Human; gene expression; heart; microarray; vascular system; probe;
XX      KW      cardiovascular disease; hypertension; cardiac arrhythmia;
XX      KW      congenital heart disease; ss.
XX      OS      Homo sapiens.
XX      PN      WO200157274-A2.
XX      PD      09-AUG-2001.
XX      PF      30-JAN-2001; 2001WO-US00666.
XX      PR      04-FEB-2000; 2000US-0180312.
XX      PR      26-MAY-2000; 2000US-0207456.
XX      PR      30-JUN-2000; 2000US-0608408.
XX      PR      03-AUG-2000; 2000US-0632366.
XX      PR      21-SEP-2000; 2000US-0234687.
XX      PR      27-SEP-2000; 2000US-0236359.
XX      PR      04-OCT-2000; 2000GB-0024263.
XX
XX      (MOLE-) MOLECULAR DYNAMICS INC.

```

```
XX Penn SG, Hanzel DK, Chen W, Rank DR:
XX WPI; 2001-48899/53.
XX Single exon nucleic acid probes for analyzing gene expression in human
XX hearts -
XX
XX Claim 4; SEQ ID No 11556; 530pp; English.
XX
XX The present invention relates to single exon nucleic acid probes for
XX measuring human gene expression in a sample derived from human heart. The
XX present sequence is one such probe. The probes may be used for
XX predicting, measuring and displaying gene expression in samples derived
XX from the human heart via microarrays. By measuring gene expression, the
XX probes are useful for predicting, diagnosing, grading, staging,
XX monitoring and prognosis diseases of the human heart and vascular system
XX e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
XX congenital heart disease.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pcl_sequences.
XX
XX Sequence 567 BP; 114 A; 124 C; 200 G; 129 T; 0 other:
XX
XX Alignment Scores:
XX Pred. No.: 1.78e-57 Length: 567
XX Score: 801.00 Matches: 151
XX Percent Similarity: 78.17% Conservative: 3
XX Best Local Similarity: 76.65% Mismatches: 3
XX Query Match: 30.95% Indels: 40
XX DB: 22 Gaps: 2
XX
US-09-902-772-4 (1-478) x ABA33090 (1-567)
QY 302 LeuAlaAsnProGlySerGlyGlnIleGlnLeuTrpGlnPheLeuLeuLeuSer 321
DB 534 CTGACCTCATTTAGGCAATGGCCAGATCCAGTTGGCACTTCCCTGGAGCTCTGCG 475
QY 322 AspSerSerAsnSerAsnCysIleThrTrpGlnGlyThrAsnGlyGluPheLeuMetThr 341
DB 474 GACAGCTCCAACTCCAGCTGCATCAGCTGGGAAGGCACCAAGGGGAGTTCAAGATGACG 415
QY 342 AspProAspGluValAlaIleGATGTGTGGLuArgIlySerIlyProAsnMetAsnTyr 361
DB 414 GATCCCAACAGGAGTGGCCGGCTGGGAGAGCGGAAGCAACCAACATGAACTAC 355
QY 362 AspIlySerAsnArgAlaIleuArgTyrTyrTrpAspIlyAsnIleMetThrIlyValHis 381
DB 354 GATAGCTCAAGCCGCGCTCCGTTACTACTATGACAAAGAACATGACCAAGTGC--- 298
QY 382 ProProGluSerSerMetTyrIlyTyrProSerAspLeuProTyrMetSerSerTyrHis 401
DB 297 -----CAT 295
QY 402 GlyIlyAsnArgTyrAlaTyrIlyPheAspPheHisGlyIleAlaGlnAlaLeuGlnPro--- 420
DB 294 GGGAAAGCGTACGCGCTCAAGTTCGACTCCACGGGATCCCGCGCTCCAGCCAC 235
QY 421 -----His 421
DB 234 CCCCCGAGTCATCTGTGCACAGTACCCCTCCAGACTCCCGGTACAGGGCTCCTATAC 175
QY 422 AlaHisProGlnIlySmeAsnPheValAlaProHisProProAlaLeuProValThrSer 441
DB 174 GCCACCCACAGAAAGTAAGTGTGTGGCGCCACCCCTCAGACCCCTCCCGTACATCT 115
QY 442 SerSerPhePheAlaAlaProAsnProTyrTyrTrpAsnSerProThrGlyIlyLeTyrPro 461
DB 114 TCCAGTTTGTGCTGCCCAAAACCCATCTACTGGAATTCACCAACTGGGGGTATATACCC 55
QY 462 AsnThrArgLeuProAlaAlaHisMetProSerHisIleuGlyThrTyrTyr 478
|||||
```

```
DB 54 AACACTAGGCTCCCGACACGCAATATGCTTCTCATCTGGGCACTTACTAC 4
RESULT 10
ID AAK14426/c
XX AAK14426 standard; DNA: 567 BP.
XX
XX AAK14426;
XX
XX 05-NOV-2001 (first entry)
XX
XX Human brain expressed single exon probe SEQ ID NO: 14417.
XX
XX Human: brain expressed exon; gene expression analysis; probe;
XX microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
XX epilepsy; cancer; ss.
XX
XX Homo sapiens.
XX
XX W0200157275-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00667.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR:
XX WPI; 2001-483446/52.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
XX brains -
XX
XX Example 4; SEQ ID NO: 14417; 650pp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX brain. They can be used to measure gene expression in brain cell samples,
XX which may enable the diagnosis and improved treatment of nervous system
XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
XX epilepsy and cancers. The present sequence is one of the probes of the
XX invention.
XX
XX Sequence 567 BP; 114 A; 124 C; 200 G; 129 T; 0 other:
XX
XX Alignment Scores:
XX Pred. No.: 1.78e-57 Length: 567
XX Score: 801.00 Matches: 151
XX Percent Similarity: 78.17% Conservative: 3
XX Best Local Similarity: 76.65% Mismatches: 3
XX Query Match: 30.95% Indels: 40
XX DB: 22 Gaps: 2
XX
US-09-902-772-4 (1-478) x AAK14426 (1-567)
QY 302 LeuAlaAsnProGlySerGlyGlnIleGlnLeuTrpGlnPheLeuLeuLeuSer 321
DB 534 CTGACCTCATTTAGGCAATGGCCAGATCCAGTTGGCACTTCCCTGGAGCTCTGCG 475
QY 322 AspSerSerAsnSerAsnCysIleThrTrpGlnGlyThrAsnGlyGluPheLeuMetThr 341
DB 474 GACAGCTCCAACTCCAGCTGCATCAGCTGGGAAGGCACCAAGGGGAGTTCAAGATGACG 415
QY 342 AspProAspGluValAlaIleGATGTGTGGLuArgIlySerIlyProAsnMetAsnTyr 361
|||||
```

```

Db 414 GATCCGAGAGGTGGCCGCGCTGGGAGACCGGAGACCAAGCAACCAATGAACTAC 355
Qy 362 AApLySLeSeArGAlAlaLeuArGTYrTYrAspLySAAnIlLeThrLySValHis 381
Db 354 GATTAAGCTCAGCGCCCTCCGTTACTACTATGACAAAGATCATGACAAAGGTC--- 298
Qy 382 ProProGluSeSerMetYrLySyrProSeArAspLeuProLyMetSeSerTYrHis 401
Db 297 -----CAT 295
Qy 402 GLyLSArGTYrAlaLYrLySPheAsPheHisGlyIleAlaGlnAlaLeuGlnPro--- 420
Db 294 GGGAAAGCGGTAGCGCTCAAGATTGAGCTTCCACAGGATGCCAGGCCCTCCAGCCCCAC 235
Qy 421 -----His 421
Db 234 CCCCCGAGTCACTCTGTACAAAGTACCCCTCAGACCTCCCGTACATGGGCTCTATAC 175
Qy 422 AlaHisProGlnLyMetAsnPhaValAlaProHisProProAlaLeuProValThrSer 441
Db 174 GCCCACCACAGAAAGATGAACTTTGTGGCGCCCGCCACCTCCAGCCCTCCGTCATCT 115
Qy 442 SerSerPhePhaAlaAlaProAsnProTYrTYrPAsnSerProThrGlyLyIleTYrPro 461
Db 114 TCCAGTTTTTTTGTCTGCCCAAAACCATACTGGAATTCAACCACTGGGGGTATATACCC 55
Qy 462 AsnThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTYrTYr 478
Db 54 AACACTAGGCTCCCGCCACAGCATATGCTCTCATCTGGGCACTTACTAC 4
RESULT 11
ID AAK40160 standard; DNA; 567 BP.
AC AAK40160;
XX
DE 06-NOV-2001 (first entry)
XX
DE Human bone marrow expressed single exon probe SEQ ID NO: 14717.
XX
KW Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX
OS Homo sapiens.
XX
FN WO200157276-A2.
XX
PD 09-AUG-2001.
XX
PE 30-JAN-2001; 2001WO-US00668.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0633366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI: 2001-488900/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human bone marrow -
XX
XX Example 4; SEQ ID NO: 14717; 658bp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow

```

```

CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC the probes of the invention.
XX
SQ Sequence 567 BP; 114 A; 124 C; 200 G; 129 T; 0 other;
Alignment Scores:
Pred. No.: 1.78e-57 Length: 567
Score: 801.00 Matches: 151
Percent Similarity: 78.17% Conservative: 3
Best Local Similarity: 76.65% Mismatches: 40
Query Match: 30.95% Indels: 2
Db: 22 Gaps: 2
US-09-902-772-4 (1-478) x AAK40160 (1-567)
Qy 302 LeuAlaAsnProGlySerGlyGlnIleGlnLeuTrpGlnPheLeuGlnLeuLeuSer 321
Db 534 CTGACCTCATTTAGCGATGGCCAGATCCAGCTTTGGCACTTCTCTGGAGCTCTGTGG 475
Qy 322 AApSerSeArSAsnSerAsnGlyIleThrTrpGluGlyThrAsnGlyLubPheLyMetThr 341
Db 474 GACAGCTCCAACTCCAGCTGCATCATCCTGGGAAGCACCAGCGGAGTTCAAGATGAGC 415
Qy 342 AspProAspGluValAlaArgArGTrpGlyGluArgLySerLySProAsnMetAsnTYr 361
Db 414 GATCCCGAGAGGTGGCCCGCGCTGGGAGAGCGGAAAGCAAGCAACCAACTGAACCTAC 355
Qy 362 AApLySLeSeArGAlAlaLeuArGTYrTYrAspLySAAnIlLeThrLySValHis 381
Db 354 GATTAAGCTCAGCGCCCTCCGTTACTACTATGACAAAGATCATGACAAAGGTC--- 298
Qy 382 ProProGluSeSerMetYrLySyrProSeArAspLeuProLyMetSeSerTYrHis 401
Db 297 -----CAT 295
Qy 402 GLyLSArGTYrAlaLYrLySPheAsPheHisGlyIleAlaGlnAlaLeuGlnPro--- 420
Db 294 GGGAAAGCGGTAGCGCTCAAGATTGAGCTTCCACAGGATGCCAGGCCCTCCAGCCCCAC 235
Qy 421 -----His 421
Db 234 CCCCCGAGTCACTCTGTACAAAGTACCCCTCAGACCTCCCGTACATGGGCTCTATAC 175
Qy 422 AlaHisProGlnLyMetAsnPhaValAlaProHisProProAlaLeuProValThrSer 441
Db 174 GCCCACCACAGAAAGATGAACTTTGTGGCGCCCGCCACCTCCAGCCCTCCGTCATCT 115
Qy 442 SerSerPhePhaAlaAlaProAsnProTYrTYrPAsnSerProThrGlyLyIleTYrPro 461
Db 114 TCCAGTTTTTTTGTCTGCCCAAAACCATACTGGAATTCAACCACTGGGGGTATATACCC 55
Qy 462 AsnThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTYrTYr 478
Db 54 AACACTAGGCTCCCGCCACAGCATATGCTCTCATCTGGGCACTTACTAC 4
RESULT 12
ID AAI20932 standard; DNA; 567 BP.
AC AAI20932;
XX
DE 12-OCT-2001 (first entry)
XX
DE Probe #10865 for gene expression analysis in human cervical cell sample.
XX
KW Probe; human; microarray; gene expression; cervical epithelial cell;
KW cervical cancer; ss.
XX
OS Homo sapiens.
XX
FN WO200157278-A2.
XX

```

```

PD 09-AUG-2001.
XX
XX 30-JAN-2001; 2001MO-US00670.
XX
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488901/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human cervical epithelial cells -
XX
XX Claim 25; SEQ ID NO 10865; 487bp; English.
XX
XX The present invention relates to human single exon nucleic acid probes
XX (SENPs). The present sequence is one such probe. The SENPs are derived
XX from human HeLa cells. The SENPs can be used to produce a single exon
XX microarray, which can be used for measuring human gene expression in a
XX sample derived from human cervical epithelial cells. By measuring gene
XX expression, the probes are therefore useful in grading and/or staging
XX of diseases of the cervix, notably cervical cancer.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 567 BP; 114 A; 124 C; 200 G; 129 T; 0 other;
XX
XX Alignment Scores:
XX Pred. No.: 1.78e-57 Length: 567
XX Score: 801.00 Matches: 151
XX Percent Similarity: 78.17% Conservative: 3
XX Best Local Similarity: 76.65% Mismatches: 3
XX Query Match: 30.95% Indels: 40
XX DB: 22 Gaps: 2
XX
US-09-902-772-4 (1-478) x AAI20932 (1-567)
QY 302 LeuAlaAsnProGlySerGlyGlnIleGlnLeuTrpGlnPheLeuGlnLeuSer 321
DB 534 CTGACCTCATTTAGGACAGTGGCCAGATCCAGCTTGGGCAAGTTCCTCGAGACTCTGTCG 475
QY 322 AspSerSerAsnSerAsnCysIleThrTrpGlnGlyThrAsnGlyGluPheLeuMetThr 341
DB 474 GACAGCTCAACTCCACCTGCTGATCAGCTGGGAAGCACAACGGGAGTCAATGACG 415
QY 342 AspProAspGluValAlaIarGATGTPGlyGluArglySerIysProAsnMetAsnTyr 361
DB 414 GATGCCACACAGAGTGGCCGCGCTGGGAGAGCGGAAGCAACCAACCATGAACCTAC 355
QY 362 AspIysLeuSerAlaGlnAlaLeuArgTyrTyrTyrAspIysAsnIleMetThrIysValHis 381
DB 354 GATAGGCTACAGCCGCGCTCCGTTACTACTATGACAAGAACATCAAGACCAAGTCC 298
QY 382 ProProGluSerSerMetTyrIysTyrProSerAspLeuProTyrMetSerSerTyrHis 401
DB 297 -----CAT 295
QY 402 GlyIysArgTyrAlaTyrIysPheAspPheHisGlyIleAlaGlnAlaLeuGlnPro--- 420
DB 294 GGGAAAGCGCTACAGCTACAGAGTTCACACGAGATCCGCCAGGCCCTCCACGCCAC 235
QY 421 -----His 421
DB 234 CCCCCGAGTCATCTGTACAAAGTACCCTCAGACCTCCCGGTACATGGGCTCTATAC 175

```

```

QY 422 AlaHisProGlnIysMetAsnPheValAlaProHisProProAlaLeuProValThrSer 441
DB 174 GCCCACCACAGAGATGAACTTGTGGCGCCACCACCTCCAGCCCTCCGCTGACATCT 115
QY 442 SerSerPhePheAlaAlaProAsnProTyrTrpAsnSerProThrGlyIleTyrPro 461
DB 114 TCCAGTTTGTGTGCGCCCAACCCATACGTAATTCACCACTGGGGGTATATACCC 55
QY 462 AsnThrArgLeuProAlaAlaHisMetProSerHisLeuGlyIleTyrTyr 478
DB 54 AACACTAGCTCCCCACGACCATATGCTTCTCATCTGGGCACTTACTAC 4
XX
XX RESULT 13
XX AAI46176/c
XX ID AAI46176 standard; DNA; 567 BP.
XX
XX AAI46176;
XX
XX 17-OCT-2001 (first entry)
XX
XX Probe #14862 used to measure gene expression in human placenta sample.
XX
XX Probe: microarray; human; placenta; antenatal diagnosis;
XX genetic disorder; ss.
XX Homo sapiens.
XX
XX WO200157272-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001MO-US00663.
XX
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-48897/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human placenta -
XX
XX Claim 25; SEQ ID NO 14862; 654bp; English.
XX
XX The present invention relates to single exon nucleic acid probes (SENPs).
XX The present sequence is one such probe. The probes are useful for
XX producing a microarray for predicting, measuring and displaying gene
XX expression in samples derived from human placenta. The probes are useful
XX for antenatal diagnosis of human genetic disorders.
XX
XX Sequence 567 BP; 114 A; 124 C; 200 G; 129 T; 0 other;
XX
XX Alignment Scores:
XX Pred. No.: 1.78e-57 Length: 567
XX Score: 801.00 Matches: 151
XX Percent Similarity: 78.17% Conservative: 3
XX Best Local Similarity: 76.65% Mismatches: 3
XX Query Match: 30.95% Indels: 40
XX DB: 22 Gaps: 2
XX
US-09-902-772-4 (1-478) x AAI46176 (1-567)
QY 302 LeuAlaAsnProGlySerGlyGlnIleGlnLeuTrpGlnPheLeuGlnLeuSer 321
DB 534 CTGACCTCATTTAGGACAGTGGCCAGATCCAGCTTGGGCAAGTTCCTCGAGACTCTGTCG 475

```

```

Db 534 CTGACCTCATTAAGCAGATGCGCAGATCCAGCTTTGGCAGTTCCTCGAGCTTCCTGCG 475
Oy 322 AspSerSerAsnSerAsnGlySerIleThrTrpGluGlyThrAsnGlyGluPheLysMetThr 341
Db 474 GACAGCTCCAACTGACCTGACCTGAGGAGCCAAACGGGGAGTTCAAGATGACG 415
Oy 342 AspProAspGluValAlaArgArgTrpGlyGluArgLysSerLysProAsnMetAsnTrp 361
Db 414 GATCCCGACGAGGTGGCCCGCCCTGGGAGAGCGGAGAGCAACCCCAACATGACATAC 355
Oy 362 AspLysLeuSerArgAlaLeuArgTrpTrpTrpAspLysAsnIleMetThrLysValHis 381
Db 354 GATTAAGCTCAGCGCGCCCTCCCTGCTACTACTATGACAGAACATCATGACCAAGTGC--- 298
Oy 382 ProProGluSerSerMetLysTrpProSerAspLeuProLysMetSerSerTrpHis 401
Db 297 -----CAT 295
Oy 402 GlyLysArgTrpAlaTrpLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnPro--- 420
Db 294 GGGAGCGCTACCGCTTACAGATTGCTGACCTTCACGAGATGCGCCAGCCCTCCAGCCCGAC 235
Oy 421 -----His 421
Db 234 CCCCCGAGTCATCTCTGTAACAAGTACCCCTCAGACCTCCCGTACATGGGCTCTATCAC 175
Oy 422 AlaHisProGlnLysMetAsnPheValAlaProHisProProAlaLeuProValThrSer 441
Db 174 GCGCCACCCACAGAAAGATGAACCTTGTGGCGCCCGCCACCTCCAGCCCTCCGAGACATCT 115
Oy 442 SerSerPhePheAlaAlaProAsnProLysTrpTrpAsnSerProThrLysGlyIleTrpPro 461
Db 114 TCCAGTTTTCCTGCTCCCAACCCATCTGGAATTCACCAACTGGGGTATATACCC 55
Oy 462 AsnThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTrpTrp 478
Db 54 AACACGAGCTCCCGCCACGACCATATGCTTCATCTGGGCGACTTACTAC 4
RESULT 14
AA106643/c
ID AA106643 standard; DNA; 567 BP.
XX
XX AA106643;
AC
XX 09-OCT-2001 (first entry)
DE Probe #6634 used to measure gene expression in human breast sample.
XX
XX Probe: human; breast disease; breast cancer; development disorder; ss;
KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
XX Homo sapiens.
OS
XX
XX WO200157270-A2.
PN
XX
XX 09-AUG-2001.
PD
XX
XX 29-JAN-2001; 2001WO-US00661.
PF
XX
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-476286/51.
XX

```

```

PT Novel single exon nucleic acid probe used to measuring gene expression
PT in a human breast -
PS Claim 25; SEQ ID NO 6634; 322bp; English.
XX
XX
CC The present invention relates to novel single exon nucleic acid probes.
CC The present sequence is one such probe. The probes are useful for
CC measuring human gene expression in a human breast sample, where the probe
CC hybridizes at high stringency to a nucleic acid expressed in the human
CC breast. The probes are useful for predicting, diagnosing, grading,
CC staging, monitoring and prognosing diseases of the human breast,
CC particularly those diseases with polygenic aetiology. The diseases
CC include: breast cancer, disorders of development, inflammatory diseases
CC of the breast, fibrocystic changes, proliferative breast disease and
CC non-carcinoma tumours.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 567 BP; 114 A; 124 C; 200 G; 129 T; 0 other;
SQ
Alignment Scores:
Pred. No.: 1,78e-57 Length: 567
Score: 801.00 Matches: 151
Percent Similarity: 78.178 Conservative: 3
Best Local Similarity: 76.658 Mismatches: 3
Query Match: 30.958 Indels: 40
DB: Gaps: 2
US-09-902-772-4 (1-478) x AA106643 (1-567)
Oy 302 LeuAlaAsnProGluSerGlyGlnIleGlnLeuTrpGlnPheLeuLeuGlnLeuSer 321
Db 534 CTGACCTCATTAAGCAGATGCGCAGATCCAGCTTTGGCAGTTCCTCGAGCTTCCTGCG 475
Oy 322 AspSerSerAsnSerAsnGlySerIleThrTrpGluGlyThrAsnGlyGluPheLysMetThr 341
Db 474 GACAGCTCCAACTGACCTGACCTGAGGAGCCAAACGGGGAGTTCAAGATGACG 415
Oy 342 AspProAspGluValAlaArgArgTrpGlyGluArgLysSerLysProAsnMetAsnTrp 361
Db 414 GATCCCGACGAGGTGGCCCGCCCTGGGAGAGCGGAGAGCAACCCCAACATGACATAC 355
Oy 362 AspLysLeuSerArgAlaLeuArgTrpTrpTrpAspLysAsnIleMetThrLysValHis 381
Db 354 GATTAAGCTCAGCGCGCCCTCCCTGCTACTACTATGACAGAACATCATGACCAAGTGC--- 298
Oy 382 ProProGluSerSerMetLysTrpProSerAspLeuProLysMetSerSerTrpHis 401
Db 297 -----CAT 295
Oy 402 GlyLysArgTrpAlaTrpLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnPro--- 420
Db 294 GGGAGCGCTACCGCTTACAGATTGCTGACCTTCACGAGATGCGCCAGCCCTCCAGCCCGAC 235
Oy 421 -----His 421
Db 234 CCCCCGAGTCATCTCTGTAACAAGTACCCCTCAGACCTCCCGTACATGGGCTCTATCAC 175
Oy 422 AlaHisProGlnLysMetAsnPheValAlaProHisProProAlaLeuProValThrSer 441
Db 174 GCGCCACCCACAGAAAGATGAACCTTGTGGCGCCCGCCACCTCCCGGAGACATCT 115
Oy 442 SerSerPhePheAlaAlaProAsnProLysTrpTrpAsnSerProThrLysGlyIleTrpPro 461
Db 114 TCCAGTTTTCCTGCTCCCAACCCATCTGGAATTCACCAACTGGGGTATATACCC 55
Oy 462 AsnThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTrpTrp 478
Db 54 AACACGAGCTCCCGCCACGACCATATGCTTCATCTGGGCGACTTACTAC 4
RESULT 15
ABS14215/c

```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 9, 2002, 17:05:02 ; Search time 55.0549 Seconds

(without alignments)
2662.642 Million cell updates/sec

Title: US-09-902-772-4

Perfect score: 2588

Sequence: 1 MASTKEALSVSEDSLF.....IYPNRLPAHMPSHLCTYY 478

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Delop 6.0 , Delext 7.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xip
-O/cgn2.1/USPTO.spool/US09902772/runat.07112002.134034.9055/app.query.fasta.1.1294
-DB=Issued_Patents_NA -OPMT=fastap -SUFFIX=rm1 -MIMATCH=0.1 -LOOPFCI=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USPR=US09902772 @cgn.1.1.40 @runat.07112002.134034.9055 -NCPU=6 -ICPU=3
-NO_XLPHY -NO_MMAP -LARGEQUERY -NEG.SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-MAIN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -YGAPOP=6 -YGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued_Patents_NA.*
1: /cgn2.6/prodata/1/ina/5A.COMB.seq.*
2: /cgn2.6/prodata/1/ina/5B.COMB.seq.*
3: /cgn2.6/prodata/1/ina/6A.COMB.seq.*
4: /cgn2.6/prodata/1/ina/6B.COMB.seq.*
5: /cgn2.6/prodata/1/ina/PCTUS.COMB.seq.*
6: /cgn2.6/prodata/1/ina/backfile1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2436	94.1	1528	4	US-08-878-177-3
2	2262.5	87.4	1447	4	US-08-878-177-1
3	1482	57.3	2938	2	US-08-343-443B-3
4	481	18.6	1752	4	US-09-360-779-1
5	481	18.6	1752	4	US-09-435-335-1
6	442.5	17.1	2268	3	US-09-344-579-1
7	440.5	17.0	1604	1	US-08-306-691B-43
8	440.5	17.0	1604	5	PCT-US93-06251-9
9	305.5	11.8	2667	2	US-08-469-412A-1
10	305.5	11.8	2667	4	US-09-021-715-1
11	301.5	11.6	328	2	US-08-343-443B-5
12	299.5	11.6	2064	3	US-08-875-944B-1

13	299.5	11.6	2064	4	US-09-116-049-3	Sequence 3, Appl1
14	297.5	11.5	1905	4	US-09-055-113-2	Sequence 2, Appl1
15	279.5	10.8	2410	2	US-08-780-835B-1	Sequence 1, Appl1
16	279.5	10.8	2410	4	US-09-303-268-1	Sequence 1, Appl1
17	279.5	10.8	2410	4	US-09-116-049-1	Sequence 1, Appl1
18	266.5	10.3	2266	2	US-09-213-767-1	Sequence 1, Appl1
19	257	9.9	5510	3	US-09-009-913-3	Sequence 3, Appl1
20	245	9.5	5427	3	US-09-009-913-2	Sequence 2, Appl1
21	245	9.5	5667	3	US-09-009-913-4	Sequence 4, Appl1
22	230	8.9	2544	2	US-08-469-412A-6	Sequence 6, Appl1
23	230	8.9	2544	4	US-09-021-715-6	Sequence 6, Appl1
24	229	8.8	848	3	US-09-009-913-38	Sequence 38, App
25	213.5	8.2	852	4	US-09-020-956-44	Sequence 44, Appl
26	213.5	8.2	852	4	US-09-030-607-44	Sequence 44, Appl
27	213.5	8.2	852	4	US-09-605-785-44	Sequence 44, Appl
28	213.5	8.2	852	4	US-09-439-313-44	Sequence 44, Appl
29	213.5	8.2	852	4	US-09-352-616A-44	Sequence 44, Appl
30	213.5	8.2	852	4	US-09-232-149A-44	Sequence 44, Appl
31	213	8.2	2280	3	US-09-009-913-8	Sequence 8, Appl1
32	213	8.2	2428	3	US-09-009-913-6	Sequence 6, Appl1
33	213	8.2	2498	3	US-09-009-913-10	Sequence 10, Appl
34	211	8.2	1920	1	US-08-746-789A-1	Sequence 1, Appl1
35	203	7.8	2975	1	US-08-368-281-1	Sequence 1, Appl1
36	199	7.7	3240	1	US-08-368-281-3	Sequence 3, Appl1
37	156	6.0	65042	4	US-09-784-316-3	Sequence 3, Appl1
38	155.5	6.0	1364	5	PCT-US93-06251-65	Sequence 65, Appl
39	155.5	6.0	1364	5	PCT-US93-06251-65	Sequence 65, Appl
40	129.5	5.0	50937	4	US-09-428-517-1	Sequence 1, Appl1
41	124	4.8	1818	1	US-08-889-402-4	Sequence 4, Appl1
42	124	4.8	1818	1	US-08-889-402-6	Sequence 6, Appl1
43	122.5	4.7	2626	1	US-08-156-020-3	Sequence 3, Appl1
44	121.5	4.7	2626	1	US-08-156-020-3	Sequence 3, Appl1
45	121.5	4.7	2626	1	US-08-156-020-7	Sequence 7, Appl1

ALIGNMENTS

RESULT 1
US-08-878-177-3
; Sequence 3, Application US/08878177
; Patent No. 6294354
; GENERAL INFORMATION:
; APPLICANT: Chugai Seiyaku Kabushiki Kaisha, 5001, Iwamoto et
; TITLE OF INVENTION: Cell Calcification Suppressing Proteins and Genes of
; FILE REFERENCE: chugai seiyaku kabushiki kaisha 5001
; CURRENT APPLICATION NUMBER: US/08/878,177
; CURRENT FILING DATE: 1997-06-18
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1528
; TYPE: DNA
; ORGANISM: C-erg gene, chicken DNA
US-08-878-177-3
Alignment Scores:
Pred. No.: 3 696-237
Score: 2436.00
Percent Similarity: 91.97%
Best Local Similarity: 91.97%
Query Match: 94.13%
DB: 4
Gaps: 2
US-09-902-772-4 (1-478) x US-08-878-177-3 (1-1528)
QY 1 MetalaserthrilytysgualaleuSERyAlValsergiuaSPGlnSerleupheGlu 20
|||||
Db 63 ATGCAAGACACTATTAGGAAGACATTACAGTGGAGATGAAGACCACTCTTTGTTGAG 122
|||||
QY 21 CysAlatryrGlyserProHlsLeuAlalyThrGluwetrAlaserSerSerGlu 40
|||||
Db 123 TGTGCTTACGATGCGCCACCTTGCAAAAGACAGAAATGACAGCCTCTTCACAGTGA 182
|||||

QY 41 TyrGlyGlnThrSerLysMetSerProArgValProGlnGlnAspTyrLeuSerGlnPro 60
 |||||
 Db 183 TATGGGAAACATCAAAAGATGAGCCGCGCTCCAGAGAGCTGTATATACACACCC 242
 QY 61 ProAlaArgValThrIleLysMetGluCysAspProAsnGlnValAsnGlySerArgAsn 80
 |||||
 Db 243 CCGGCCAGAGTTACCTTAAGATGAGAGTGTAAACCAACAGGTTAATGGGTCAAGAAAT 302
 QY 81 SerProAspAspCysSerValAlaLysGlyLysMetValSerSerSerAspAsnVal 100
 |||||
 Db 303 TCACCTGATGACTGACAGCGGCGCAAAAGAGGAAAAATGGTTAGCAGTCAAAATGTT 362
 QY 101 GlyMetAsnTyrGlySerTyrMetGluGluLysHisIleProProProAsnMetThrThr 120
 |||||
 Db 363 GGGATGAACATATGAAAGCTACATGAGAGAGAACATATTCGCTCCAAATATGACAAAC 422
 QY 121 AsnGluArgArgValIleValProAlaAspProThrLeuTyrSerThrAspHisValArg 140
 |||||
 Db 423 AATGAAGCAAGAGATTTATGTGGCAGAGATCCATACGTTATGAGACACAGACATGTACGG 482
 QY 141 GlnTyrLeuGluTyrPalaValLysGluTyrGlyLeuProAspValAspIleLeuLeuPhe 160
 |||||
 Db 483 CAGTGGCTGGAGTGGCGAGAGAGAGATATGGCTTCACAGCTGAGACATCTTGTGTTTC 542
 QY 161 GlnAsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArgLeuThr 180
 |||||
 Db 543 CAGAACATTTGATGGAAAGAGTGTATAAATACCAAGAGTACTCCAGAACACTCAGC 602
 QY 181 ProSerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGluThrProLeu 200
 |||||
 Db 603 CCGAGTATACGACATATTCCTCTGTCACACTACACTACAGAGAGATCCTCTT 662
 QY 201 ProHisLeuThrSerAspAspValAspLysAlaLeuGlnAsnSerProArgLeuMetHis 220
 |||||
 Db 663 CCACATTTACTCTACATGATGTGTAAAGCCCTTACAAACCTCCACGCTTAATGCAT 722
 QY 221 AlaArgAsnThrGlyLysIleThrPheIlePheProAsnThrSerAlaTyrProGluAla 240
 |||||
 Db 723 GCTAGAAACACAGGAGAGACCACTTTATTTTCCAAATACATCAGATTAAACCAAGAGCA 782
 QY 241 ThrGlnArgIleThrThrArgProAspLeuProTyrGluGlnAlaArgArgSerAlaTyr 260
 |||||
 Db 783 AGCAAGATATTAACAACAAGCCAGATTACCTTATGACCAAGCAGAGATACGGCTGG 842
 QY 261 ThrSerHisSerHisProThrGlnSerLysAlaThrGlnProSerSerThrValPro 280
 |||||
 Db 843 ACGAGTCACAGCCATCCACTCAGTCAAAAGCTAACCAACATCATCTTCAACAGTGGCC 902
 QY 281 LysThrGluAspGlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSerSer 300
 |||||
 Db 903 AANAACGAAGACCCAGCGTCTCAGTTAGATCCCTTATCAAGATTCCTTGGACCGACACAGC 962
 QY 301 ArgLeuAlaAsnProGlySerGlyGlnIleGlnLeuTyrGlnPheLeuLeuGluLeuLeu 320
 |||||
 Db 963 CGCTTGCAGAAATCCAGGGAGTGGCGAGATACAGCTATGCGAGTCTCTACTGGAGCTTCG 1022
 QY 321 SerAspSerSerAsnSerAsnCysIleThrTyrGluGlyThrAsnGlyLysPheLysMet 340
 |||||
 Db 1023 TCGGACAGTCCAACTCCAACTGCAATCAGCTGGAGGGCAAAATGGGAGTTCAAGATG 1082
 QY 341 ThrAspProAspGluValAlaArgArgTyrGlyGluArgLysSerLysProAsnMetAsn 360
 |||||
 Db 1083 AAGAGACCTGATGAAAGTGGCTGGCTGGGAGAGAGAGAAAAAGCAAAACCTTAACATGAAC 1142
 QY 361 TyrAspLysLeuSerArgAlaLeuArgTyrTyrTyrAspLysAsnIleMetThrLysVal 380
 |||||
 Db 1143 TATGACAAAACTCAGCGCTGCACTTGGCTACTACTATGACAAAAATATTTATGACTAAAGTT 1202
 QY 380 ----- 380
 Db 1203 CATGGTAAACGCTATGCTACAAATTTGATTTCCACGGAATGCTTCAGGCTCCAGCCT 1262

QY 381 HisProGluSerSerMetTyrLysTyrProSerAspLeuProTyrMetSerSerTyr 400
 |||||
 Db 1263 CACCCTTCAGAAATCATCAATGATCAAAATATACCATCAACATCCCTTACATGAGTCTTAC 1322
 QY 401 HisGlyLysArgTyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnPro 420
 |||||
 Db 1322 ----- 1322
 QY 421 HisAlaHisProGluLysMetAsnPheValAlaProHisProProAlaLeuProValThr 440
 |||||
 Db 1323 CATGCAACACCCCGAAGAGATGACTTGTAGCTCCCATCCCTGCTTGGCCGTTAAC 1382
 QY 441 SerSerSerPhePheAlaAlaProAsnProTyrTyrPasnSerProThrGlyGlyIleTyr 460
 |||||
 Db 1383 TCATCCAGCTTTTGTGCTGCCCTAATCCATGATGGAAATTCACCAACTGGAGGCATCTAC 1442
 QY 461 ProAsnThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr 478
 |||||
 Db 1443 CCCAATRACAGGCTGCCAGCTGCTCATATGCCCTTCCCATCTTGGACACTRATAC 1496
 RESULT 2
 US-08-878-177-1
 ; Sequence 1, Application us/08878177
 ; Patent No. 6294354
 ; GENERAL INFORMATION:
 ; APPLICANT: Chugai Selyaku Kabushiki Kaisha, 5001, Iwamoto et
 ; TITLE OF INVENTION: Cell Calcification Suppressing Proteins and Genes of
 ; TITLE OF INVENTION: the proteins
 ; FILE REFERENCE: Chugai Selyaku Kabushiki Kaisha 5001
 ; CURRENT APPLICATION NUMBER: US/08/878,177
 ; NUMBER OF SEQ ID NOS: 7
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 1
 ; LENGTH: 1447
 ; TYPE: DNA
 ; ORGANISM: C-11 gene, c-ery gene w/ deletion, chicken DNA
 US-08-878-177-1
 Alignment Scores:
 Pred. No.: 1,28e-219 Length: 1447
 Score: 2262,50 Matches: 430
 Percent Similarity: 86,35% Conservative: 0
 Best Local Similarity: 86,35% Mismatches: 1
 Query Match: 87,42% Indels: 67
 Ds: 4 Gaps: 3
 US-09-902-772-4 (1-478) x US-08-878-177-1 (1-1447)
 QY 1 MetaLaserThrIleLysGluAlaLeuSerValSerGluAspGlnSerLeuPheGlu 20
 |||||
 Db 63 ATGGCAAGACATATTAAAGAGAGCATATACATGATGATGAGAGACAGACATCTTGTGAG 122
 QY 21 CysAlaTyrGlySerProHisLeuAlaLysThrGlnMetThrAlaSerSerSerSerGlu 40
 |||||
 Db 123 TGTGCTTACGATCCGCCCACTCTGCAGAAACAGAAATACAGCCCTCTTCCAGTGA 182
 QY 41 TyrGlyGlnThrSerLysMetSerProArgValProGlnGlnAspTyrLeuSerGlnPro 60
 |||||
 Db 183 TATGGCAAAATCAAAAGATGAGCCCGCGCTTCCACAGAGACTGTTATCACAGCCC 242
 QY 61 ProAlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsn 80
 |||||
 Db 243 CCGGCCAGAGTTACATTTAAGATGAGTGTAAACCAACCAAGATTATGGTCAAGAAAT 302
 QY 81 SerProAspAspCysSerValAlaLysGlyLysMetValSerSerSerAspAsnVal 100
 |||||
 Db 303 TCACCTGATGACTGACGCGTGGCAAAAGAGGAAAAATGTATGACAGATTCATGATGTT 362
 QY 101 GlyMetAsnTyrGlySerTyrMetGluGluLysHisIleProProProAsnMetThrThr 120
 |||||
 Db 363 GGGATGAACATATGAAAGCTACATGAGAGAGAACATATTCGCCCTTCCAAATATGACAAAC 422

OY	121	ASGIuAgtAgtValIIEvalProAlaAspProThleuTrpSerThrAspHisValaG	140
OY	121	ASGIuAgtAgtValIIEvalProAlaAspProThleuTrpSerThrAspHisValaG	140
Db	423	AATGACAGAAAGATTATGTGGCCAGCAGATCTCACTTATGGAGCACAACAATGTACGG	482
OY	141	GIuTrpLeuGIuTrpAlaValaValysGIuTyGIuLeuProAspValaAspIleLeuPhe	160
Db	483	CAGTGGCTGGAGTGGGCGAGTAAGAGATATGTCCTCCAGACCTGGACATCTTGTTGTC	542
OY	161	GIAspIIEAspGIuValysGIuLeuCysLysMetThrLysAspAspPheGIuArgLeuThr	180
Db	543	CAGAACATTGATGGGAAAGAGTGTGTAAATGATACCAAAAGATGACTCTCAGAGACTCAG	602
OY	181	ProSerThrAsnAlaAspIleLeuLeuSerHisLeuHisIstyreuArgGIuThrProLeu	200
Db	603	CCGAGCTATTAACGGAGATATCTCTCTGTCACACCTCACTCACTCGAGAGAGGA-----	656
OY	201	ProHisLeuThrSerAspAspValaAspLysAlaLeuGIuAsnSerProArgLeuMetHis	220
Db	656	-----	656
OY	221	AlaArgAsnThrGIuGIuAlaThrPheIlePheProAsnThrSerValaTyuProGIuAla	240
Db	657	-----GGAGCCACTTTTATTTTCCAAATATCAATCACTTATACCAGAGA	701
OY	241	ThrGIuArgIleThrThrArgProAspLeuProTyGIuGIuAlaArgAspSerAlaTrp	260
Db	702	ACGCAAAAGAAATACMACAGAGCCAGATTTACCTTATAGACMACCGAGGAGATCAGCTGG	761
OY	261	ThrSerHisSerHisProThrGIuInsLeuLysAlaThrGIuProSerSerSerThrValPro	280
Db	762	ACGAGTCAACAGCCATCCCACTCACTCACTCAAAAGCTAACCCACATATCTTCAACAGTGCC	821
OY	281	LysThrGIuAspGIuArgProGIuInsLeuAspProTyGIuIIEuGIuIProThrSerSer	300
Db	822	AAAACAGAAAGCAGCGCTCCTCAGATTGATCTTACAGATCTTGAGACCGACAGCAGC	881
OY	301	ArgLeuAlaAsnProGIuSerGIuGIuIIEGIuLeuTrpGIuPheLeuLeuGIuLeuLeu	320
Db	882	CGTTCCTCAATTCAGAGGAGATGGCGAGATACACCTTATGGCAGATCTCTATGGAGCTTCG	941
OY	321	SerAspSerSerAsnSerAsnCysIIEThrTrpGIuGIuThrAsnGIuIuPheLysMet	340
Db	942	TTCGACAGCTCCAACTCCAACTGCATCACCTGGAGAGGCAAAATGGAGTTCAAGATG	1001
OY	341	ThrAspProAspGIuValaAlaArgArgTrpGIuIArgLysSerLysAspProAsnMetAsn	360
Db	1002	ACAGACCTGATGAAGTGGCTGGCGTGGGGAGAGAGAAAGCAAACTTAACATGAAC	1061
OY	361	TyrAspLysLeuSerArgAlaLeuArgTyuTyuTyuAspLysAsnIIEuThrLysVal	380
Db	1062	TATGACAAATCAGCCCTGCATCTCGTCTACTATGACAAAAATATGTATGACTTAAGTT	1121
OY	380	-----	380
Db	1122	CATGGTAAACGGCTATGGCTACAAATTTGATTCTCCACGGAATCGCTCAGGCCCTCAGCCT	1181
OY	381	HisProProGIuInsSerSerMetTyuTyuTyuProSerAspLeuProTyuMetSerSerTyu	400
Db	1182	CACCTCCAGATATATCCATGTCAAATATCCATACGACTCCCTCAGTGAATGATCTTCAC	1241
OY	401	HisGIuLysArgTyuAlaTyuTyuLysPheAspPheHisGIuIIEalGIuAlaLeuGIuPro	420
Db	1241	-----	1241
OY	421	HisAlaHisProGIuInsMetAsnPheValAlaProHisProProAlaLeuProValaThr	440
Db	1242	CATGCAACACCCCGAAGAGAACTTTGTAAGCTCCCAATCCCCCTGCTTTGGCCCTAACCC	1301
OY	441	SerSerSerPhePheAlaAlaProAsnProTyuTrpAsnSerProThrGIuIIEuTyuLeuTyu	460
Db	1302	TCAATCCAGCTTTTGTGTGCCCTATATCCATATCGAATATTCACCAACTGGAGGCACTTACC	1361
OY	461	ProAsnThrArgLeuProAlaAlaHisMetProSerHisLeuGIuIuThrTyuTyu 478	

```

Db      1362  CCCAATACAGGCTGCACGCTGCTCATATGCGCTTCCCATCTTGGCAGCTACTACTAC 1415
|||||
RESULT 3
US-08-343-443B-3
: Sequence 3, Application US/08343443B
: Patent No. 5968734
: GENERAL INFORMATION:
: APPLICANT: Aurias, Alain
: APPLICANT: Delattre, Olivier
: APPLICANT: Desmaze, Chantal
: APPLICANT: Melot, Thomas
: APPLICANT: Peter, Martine
: APPLICANT: Ploougaestel, Beatrice
: APPLICANT: Thomas, Gilles
: APPLICANT: Zucman, Jessica
: TITLE OF INVENTION: NUCLEIC ACID CORRESPONDING TO A GENE OF
: TITLE OF INVENTION: CHROMOSOME 22 INVOLVED IN RECURRENT CHROMOSOMAL
: TITLE OF INVENTION: TRANSLATIONS ASSOCIATED WITH THE DEVELOPMENT OF CANCEROUS
: TITLE OF INVENTION: TUMORS, AND NUCLEIC ACIDS OF FUSION RESULTING FROM SAID
: TRANSLATIONS
: NUMBER OF SEQUENCES: 129
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Weiser & Associates
: STREET: 230 South Fifteenth Street
: CITY: Philadelphia
: STATE: PA
: COUNTRY: USA
: ZIP: 19102
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: AEDIT 1.0 DOS text editor
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/343,443B
: FILING DATE: 18-NOV-1994
: CLASSIFICATION: 514
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/FR93/00494
: FILING DATE: 19-MAY-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: FR 92/06123
: FILING DATE: 20-MAY-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Weiser, Gerard J.
: REGISTRATION NUMBER: 19,763
: REFERENCE/DOCKET NUMBER: 989.6121P
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 215-875-8383
: TELEFAX: 215-875-8394
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2938 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 143..1498
US-08-343-443B-3

Alignment Scores:
Pred. No.:      4,88e-140      Length:      2938
Score:          1482.00      Matches:      291
Percent Similarity: 68.65%      Conservative: 55
Best Local Similarity: 57.74%      Mismatches:  80
Query Match:      57.26%      Indels:      78
DB:                2          Gaps:                9
US-09-902-772-4 (1-478) x US-08-343-443B-3 (1-2938)
1 MetalaserThrIlelysglnAlaIeuSerValIaSerGlnIuaspGlnSerIeuPheGlu

```

```

Db 143 ATGGACGGGAGCTATTGAGAGGCTGCTGCTGCTGAGGAGGACGACGACCTTTTGAC 202
Oy 21 CysAlaTyrGly---SerProHisLeuAlaLysThrGluMetThrAlaSerSerSer 39
Db 203 TCAGCGTACGAGGAGCGGACGCCATCTCCCGCAGGCCGACATGATGCTCTCGGGAGTCT 262
Oy 40 GluTyrGlyGlnThrSerLysMetSerProArgValProGlnGlnAspTrpLeuSerGln 59
Db 263 GACTAGCGGACAGCCGACAAAGATCAACCCCTCCACACAGAGAGTGGATCAATCAG 322
Oy 60 ProProAlaArgValThrLleLysMetGluCysAsnProAsnGlnValAsnGlySerArg 79
Db 323 ---CCAGTGGAGGTCACCTCAAGCGGAGTAT---GACCACATGATGATTCACAG 373
Oy 80 AsnSerProAspAspCysSerSerValAlaLysGlyGlyLysMetValSerSerSerAsp 99
Db 374 GAGTCTCGCGTGGACTGCGAGCGCTTACAAATGACAGCAGCTGTGGCGGAGCGGAGTCC 433
Oy 100 ValGlyMetAsnTyrGlySerTyrMetGluGluLysHis---LleProProAsnMet 118
Db 434 AACCCCATGACTACACACAGCTATATGACGAGAAAGATGCCCTCCCTCCCAACATG 493
Oy 119 ThrThrAsnGlnArgArgValLleValProAlaAspProThrLeuTrpSerThrAspHis 138
Db 494 ACCACCAACGAGAGGAGTCAATCGTCCCGCAGACCCACACTGTGGACAGGAGCAT 553
Oy 139 ValArgGlnTrpLeuGluTrpAlaValLysGluTyrGlyLysProAspValAspLleu 158
Db 554 GTGAGGCAATGGCTGGAGTGGGCGCATAAAGAGTATAGCTTATGATGAGATCCACATCC 613
Oy 159 LeuPheGlnAsnLleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArg 178
Db 614 TTTTTCAGAACATGATGATGCGCAAGACAGTGTAAATCAACAGAGGAGCTTCCCTCC 673
Oy 179 LeuThrProSerTyrAsnAlaAspLleLeuLeuSerHisLeuHisTyrLeuArgGluThr 198
Db 674 GCCACACCCCTTACACACGAGAGTGGTGTGTCAACCTGATACCTGATTCAGGAGAAAT 733
Oy 199 ProLeuProHisLleuThrSerAspAspValAspLysAlaLleuGlnAsnSerProArgLeu 218
Db 734 TCACGTG----- 739
Oy 219 MethAlaLysArgAsnThrGlyGlyAlaThrPheLlePheProAsnThrSerValTyrPro 238
Db 740 -----CTGGCCTATAATAACACCTCCACACACC 766
Oy 239 GluAlaThrGlnArgLleThrArgProAspLeuProTyrGluGlnAlaArgArgSer 258
Db 767 GACCAATCTCTACGATGTGCTCAAGAGAACCTTCTTATGACTCAGTCAAGAGAGGA 826
Oy 259 AlaTrpThrSerHisSerHisProThrGlnSerLysAlaThrGln---ProSerSerSer 277
Db 827 GCATGGGGCAATACATGATTCGTGGCTCAACAAAGTCCCTCCCTTGGAGGGGCACAA 886
Oy 278 ThrValProLysThrGluAspGlnArgProGlnLeuAspProTyrGlnLleLeuGlyPro 297
Db 887 ACGATCAGTAAATATACAGAGCAACGCGCCCGCCAGATCCCTATCAGATCTCGGGCCG 946
Oy 298 ThrSerSerArgLeuAlaAsnProGlySerGlyGlnLleGlnLeuTrpGlnPheLeu 317
Db 947 ACCACAGTGGCTTACCAACCTGAGAGGGGCGGAGATCCAGCTGTGGCAATTCCTCTCG 1006
Oy 318 GluLeuLeuSerAspSerSerAsnSerAsnCysLleThrTrpGluGlyThrAsnGlyGlu 337
Db 1007 GAGTGTCTCTCGACAGGCGCAACGCCAGCTATCACCCTGGAGGAGGACCAACGGGAG 1066
Oy 338 PheLysMetThrAspProAspGluValAlaArgArgTyrGlyGlnArgLysSerLysPro 357
Db 1067 TTCAAAATATACAGGAGCCGATAGGTGGCCAGAGGCTGGGGCGAGCGGAAAGCAAGCCC 1126
Oy 358 AsnMetAsnTyrAspLysLeuSerArgAlaLeuArgTyrTyrTyrAspLysAsnLleMet 377

```

```

Db 1127 AACATGAATTACGACAACTGAGCGGGCCCTCCGTTATTACTANGATAAAAATTATG 1186
Oy 378 ThrLysValHisProProGluSerMetTyrLysTyr-ProSerAspLeuProTyrMet 397
Db 1187 ACCAAAGTC----- 1195
Oy 398 SerSerTyrHisGlyLysArgTyrAlaTyrLysPheAspPheHisGlyLleAlaGlnAla 417
Db 1196 -----CACGGCAAAAGATATGCTTCAAAATTTGACTTCCACGGCATTCGCCAGCT 1246
Oy 418 LeuGlnPro----- 420
Db 1247 CTGCACGACCAATCCGACGAGCTGCCATGTACAAAGTACCTTTCAGACTCTACATG 1306
Oy 421 -----HisAlaHisProGlnLysMetAsnPheValAlaProHisProProAlaLeu 437
Db 1307 CCTTCCTACCAATGCCACGACGAGAAAGTGAATTTGTCCCTCCCATTCATCCCTCAG 1366
Oy 438 ProValThrSerSerSerPhePheAlaAlaProAsnProTyrTrpAsnSerProThrGly 457
Db 1367 CCGTACACTTCCCTCCAGCTTCTTGGAGCGCGATCATCAATACTGGACACTCCCGG 1426
Oy 458 GlyLleTyrProAsn-----ThrArgLeuProAlaAlaHisMetProSerHisLeu 474
Db 1427 GGAATCTACCCCAACCCCAACGTCGCCGCACTCAACACCCACGCTGCTTACACTTA 1486
Oy 475 GlyThrTyrTyr 478
Db 1487 GGCAGCTACTAC 1498

RESULT 4
US-09-360-779-1
; Sequence 1, Application US/09360779
; Patent No. 6268216
; GENERAL INFORMATION:
; APPLICANT: Denertis, Evan S.
; APPLICANT: Ryodor, Dmitry V.
; APPLICANT: Hendricks, Timothy J.
; TITLE OF INVENTION: Reagents and Methods for the Screening of Compounds
; TITLE OF INVENTION: Useful in the Treatment of Neurological Diseases
; FILE REFERENCE: CASE-03828
; CURRENT APPLICATION NUMBER: US/09/360,779
; CURRENT FILING DATE: 1999-07-26
; EARLIER APPLICATION NUMBER: 60/094,264
; EARLIER FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1752
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (112)..(1131)
US-09-360-779-1

Alignment Scores:
Pred. No.: 5.54e-39 Length: 1752
Score: 481.00 Matches: 122
Percent Similarity: 44.73% Conservative: 18
Best Local Similarity: 38.98% Mismatches: 68
Query Match: 18.59% Indels: 105
DB: 4 Gaps: 10

US-09-902-772-4 (1-478) x US-09-360-779-1 (1-1752)
Oy 224 ThrGlyGlyAlaThrPheLlePheProAsnThrSerValTyrProGlnAlaThrGlnArg 243
Db 333 ACGGCGGGGTATCGGACCCCAACGCGCAAGCTGACGTGCCCCCGGTGGCTCCCGCC 392
Oy 244 IleThrThrArgProAspLeuProTyrGluGlnAlaArgArgSerAlaTrpThrSerHis 263
Db 393 ATCTCCACCGCGCA-----GTCCCGGCG 416

```


DB 996 CTCT-----TACTGGCTGCTCCAAAGCCACCGCCGCTGC 1031
OY 456 -----ThrglyGlyIleTyrProAsnThrArgLeuPro----- 466
DB 1032 CGCGCCGACCGGCGCTCTACCAACCGCGGCTTGCAAGCCGCCCTTGG 1091
OY 467 -AlaAlaHisMetProSerHisLeuGlyThrTyr 478
DB 1092 CGGGGTGGCCGCGCTTGCACCTTGGGGGTCAATTAT 1128
RESULT 6
US-09-344-579-1
; Sequence 1, Application US/09344579
; Patent No. 6054316
; GENERAL INFORMATION:
; APPLICANT: Brenda F. Baker
; APPLICANT: Lex M. Cowsett
; TITLE OF INVENTION: ANTISENSE MODULATION OF EMS-2 EXPRESSION
; FILE REFERENCE: RTS-0063
; CURRENT APPLICATION NUMBER: US/09/344,579
; CURRENT FILING DATE: 1999-06-25
; NUMBER OF SEQ ID NOS: 47
; SEQ ID NO 1
; LENGTH: 2268
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (291)..(1700)
US-09-344-579-1
Alignment Scores:
Pred. No.: 6,71e-35 Length: 2268
Score: 442.50 Matches: 133
Percent Similarity: 39.50% Conservative: 40
Best Local Similarity: 30.37% Mismatches: 108
Query Match: 17.10% Gaps: 16
DB: 3 Indels: 16
US-09-902-772-4 (1-478) x US-09-344-579-1 (1-2268)
OY 122 GUATGATGValIleValProAlaAspProThrLeuTrpSerThrAspHisValArgGln 141
DB 570 CAGCGGCGCTGCGGCAATTCAGAAAGCCCTGGCTGTGAGTACAGACAGGATGCGCAG 629
OY 142 TTPLeuGluTrpAlaValAlaGlyLeuTyrGlyLeuProAspValAspIleLeuLeuPheGln 161
DB 630 TGGCTCTCTGGCCGACCAATGATGATTCAGTCTGTGCAAGTCAATTCGACAGGCTTC 686
OY 162 AsnIleAspGlyLysGluLeuLeuGlyMetThrLysAspAspPheGlnArgLeuThrPro 181
DB 687 GGCATGAATGGCCAGATGCTGTGTAACCTTGGCAAGAACGCTTCTGTGAGCTGGCAGCT 746
OY 182 SerTyrAsnAlaAspIleLeuLeuSerHisLeu----- 192
DB 747 GACTTGTGGGTGACATCTCTGTGGAACTGTGGAGCAAAATGATCAAGAAAAACAAGAA 806
OY 193 -----HisTyrIleuArg 196
DB 807 AAGACAGAGATCAATATGAGAAAAATTTCACACCTCAGCTCCGTTCTCTGATTGATTAC 866
OY 197 GluThr----- 198
DB 867 AGCAATACATTAGTTTGGACAGAGAGCGCCCTATGATGACAGACAGAAATTAC 926
OY 199 -----ProLeu-ProHisIleuThrSerAspAsp 208
DB 927 CCCAAGCGGCGCTCTGACAGACATGTGTCCGCGCTCCACACCCAGCGTACTCAGCTCT 986
OY 208 LaspYsalAlaLeuAsnSerProArgLeu----- 218
DB 987 GAGCAGAGATTCACAGATGTCTCCCAAGTCTCGGCTCAGCTCCGTCAGGCTCAGCTACTGC 1046

OY 219 -MetHisAlaArgAsnThrGlyValAlaThrPheIle-----Phe 231
DB 1047 TCTGTACGTACGAGACTTCCAGGACAGCAACTGAATTCTCTACCAACAATTCTGGAGCT 1106
OY 231 eProAsnThrSerValTyrProGlnAlaThrGlnArgIleThrThrArgProAspLeu 251
DB 1107 CCCAAGACCACGACT-----CCCTGAGAACGGTGGGACAGCTTGCAGAGCTCAGACTCC 1163
OY 251 oTyrGluGln-----AlaArgSerAlaTrp----- 260
DB 1164 CTCTCCAGTCCGGAACAGCCAGTCTCTCTCTGATGTGCAAGGGTCTCTCTTC 1223
OY 261 -----ThrSerHisSerHisProThrGlnSe 269
DB 1224 GAGAGCTTGAGAGTACTGACAGCCAGTCTCTCTGCTCAATTAAGCAACATGCTTTC 1283
OY 269 rlySalAlaThrGlnProSerSerSerThrValProLysThrGluAspGlnArgProGlnSe 289
DB 1284 AAGGATTACA-----TCCAGAGAGAGAGTACCCAGTG- GAGCAAGCAACACAGTTAT 1336
OY 289 uAspProTyrGlnIleLeuGlyProThrSerSerArgLeuAlaAsnProGlySerGly 309
DB 1337 ACCTGCAGCTGTGCTGGCGGCTTCACA-----GGAAGTGACC 1375
OY 309 nIleGlnLeuTyrPglInPheLeuLeuGluLeuSerAspSerSerAsnSerAsnCysII 329
DB 1376 TATTCAGTGTGAGACTTCTCTGAGCTGCTATCAGACAAATCTCGCAGCATTTAT 1435
OY 329 eMhrrTPGluGlyThrAsnGlyLuhPheLysMetThrAspProAspGluValAlaArg 349
DB 1436 CAGCTGACTGGAGCGAGTGGAGTTAAGCTCGCCGACCCGATGAGAGTGGCCGCG 1495
OY 349 gTPGlyGluArgLysSerLysProAsnMetAsnTyrAspLysLeuSerArgAlaLeu 369
DB 1496 GTGGGAAAGAGAAAAATTAAGCCCAAGTGAATCAAGAGCTGAGCCGGGCTTAC 1555
OY 369 gTyrTyrTyrAspLysAsnIleMetThrLysValHisProGluSerSerMetTyrLys 389
DB 1556 CTACTATTACGACAGAACATCATCCACAAGACG----- 1589
OY 389 sTyrProSerAspLeuProTyrMetSerSerTyrHisGlyLysArgTyrAlaTyrLys 409
DB 1590 -----TCGGGGAAGCGCTTACGTCGCTT 1615
OY 409 e-----AspPheHisGlyIleAlaG 416
DB 1616 CGGTGCGACCTCCAGAACTTCTGCGGTTCACAGCCCGGACAGACTGCGCATCTG 1673
OY 416 nAlaLeuGlnProHisAla-----HisProGlnLysMetAsnPh 429
DB 1674 -GGCGTCCAGCCGACAGCAGGAGACTGAGTGCAGGACCACTGAGCCG----- 1724
OY 429 eValAlaPro-----HisProProAlaLeuPro 438
DB 1725 -----GCCCAAGCGCTGTGAGTGTGGAAGCCCATCTCTGACAGCTGCT 1772
RESULT 7
US-08-306-691B-43
; Sequence 43, Application US/08306691B
; Patent No. 5734039
; GENERAL INFORMATION:
; APPLICANT: Calabretta, Bruno
; APPLICANT: Skorski, Tomasz
; TITLE OF INVENTION: ANTISENSE
; TITLE OF INVENTION: OLIGONUCLEOTIDES TARGETING COOPERATING ONCOGENES
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESS: Seidel, Gonda, Lavorgna & Monaco, P. C.
; STREET: Two Penn Center, Suite 1800
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: U.S.A.

Alignment Scores:

Pred. No.: 6.64e-21 Length: 2667
 Score: 305.50 Matches: 88
 Percent Similarity: 50.91% Conservative: 24
 Best Local Similarity: 40.00% Mismatches: 64
 Query Match: 11.80% Indels: 47
 DB: 2 Gaps: 7

US-09-902-772-4 (1-478) x US-08-469-412A-1 (1-2667)

QY 266 ProthrglnSerLysAlaThrGlnProSerSerThrValProLysThrGluAsp--G 285
 DB 74 CCGGAGCCCGGGCCCGGAATCGGGCGCTCGCCCGCGCCCGCCAGCATGAAGACCC 133
 QY 285 lnaArgProGlnLeuAspProTyrglnlleuGlnProThrsSerArgLeuAlaAsp 305
 DB 134 GCGGACACAGAGGTTGGCTTCGCGGAT---GGGCTTACA--AGCCAGAG-TGCTCC 186
 QY 305 roGlySerGlyGlnlleGlnLeuTrpGlnPheLeuLeuGlnLeuSerAspSerSera 325
 DB 187 CTGGCTCAAGGACGATCCAGCTGTGCACTTATCTGGAGCTGCTGGGAAGAGAGAGT 246
 QY 325 snSerAsnCysIleThrTrpGlnGlyThrAsnGlyGluPheLysMetThrAspProAspG 345
 DB 247 ACCAGGCGCTATTCCTGGCAGGGGAGGAGCTACGGGAATTCGATCAAGACCTGATG 306
 QY 345 luValAlaArgArgTrpGlyGluArgLysSerLysProAsnMetAsnTyraAspLysLeuS 365
 DB 307 AGTGAGCCCGCGCTGTGGGCGCTGCAAGTGCAAGCCCAAGATGAATGACACAAGCTGA 366
 QY 365 eraArgAlaLeuArgTyrrTyrrAspLysAsnIleMetThrLysValHisProGluS 385
 DB 367 GCCGGGCCCTGGCTATTAATAACAAGCGCATTCGTCACAGAAC-- 413
 QY 385 erSerMetTyrrLysTyrrProSerAspLeuProTyrrMetSerSerTyrrHisGlyLysArgT 405
 DB 414 -----AAGGGGAAACGCT 426
 QY 405 yrrAlaTyrrLysPheAspPheHisGlylleAlaGlnAlaLeuGlnProHis----- 421
 DB 427 TCACCTCAAGATTCATTCACAAACAGTGTGCTGATCAATTCATTCATTCATGATG 486
 QY 422 -----AlaHisProGlnLysMetAsnPheValAlaProHisProAlaL 437
 DB 487 GGTGGCTGGGGGTGAGTGCCTGCCAGAGT-----GCCCGCGCAG 525
 QY 437 euProValThrSerSerSerPhePheAlaAlaProAsnProTyrrTripaSerProThrG 457
 DB 526 TGCCGTGGGGGTAGCCACTTCGCTTCCTCC-----TCAAGGCCCTCG 573
 QY 457 lyGlyIleTyrrProAsnThrArgLeuProAlaAlaHisMetProSer---HisLeu 474
 DB 574 A-GGTGCTGTCGCCACCG-AGGACCCCGGTCACACACAGCAGCTGCTTCATCTT 627

RESULT 10

US-09-021-715-1
 ; Sequence 1, Application US/09021715
 ; Patent No. 6194547

GENERAL INFORMATION:
 APPLICANT: MayoThalassemia, George J.
 Blair, Donald G.
 Fisher, Robert J.
 Beal Jr., Gregory J.
 Athanasios, Meropi A.
 Sgouras, Dionysios N.

TITLE OF INVENTION: The ERF Genetic Locus and Its Products

NUMBER OF SEQUENCES: 16

CORRESPONDENCE ADDRESS:
 ADDRESS: Townsend and Townsend and Crew LLP
 STREET: Two Embarcadero Center, Eighth Floor
 CITY: San Francisco
 STATE: California
 COUNTRY: USA

ZIP: 94111-3834
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/021,715
 FILING DATE: 10-Feb-1998
 CLASSIFICATION: <Unknown>
 ATTORNEY/AGENT INFORMATION:
 NAME: Garrett-Mackowski, Eugenia
 REGISTRATION NUMBER: 37,330
 REFERENCE/DOCKET NUMBER: 015280-229000
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 576-0200
 TELEFAX: (415) 576-0300
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2667 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 123..1769
 OTHER INFORMATION: /note= "human ERF (ERS2 Repressor
 Factor) cDNA"
 SEQUENCE DESCRIPTION: SEQ ID NO: 1:
 US-09-021-715-1

Alignment Scores:

Pred. No.: 6.64e-21 Length: 2667
 Score: 305.50 Matches: 88
 Percent Similarity: 50.91% Conservative: 24
 Best Local Similarity: 40.00% Mismatches: 64
 Query Match: 11.80% Indels: 47
 DB: 4 Gaps: 7

US-09-902-772-4 (1-478) x US-09-021-715-1 (1-2667)

QY 266 ProthrglnSerLysAlaThrGlnProSerSerThrValProLysThrGluAsp--G 285
 DB 74 CCGGAGCCCGGGCCCGGAATCGGGCGCTCGCCCGCGCCCGCCAGCATGAAGACCC 133
 QY 285 lnaArgProGlnLeuAspProTyrglnlleuGlnProThrsSerArgLeuAlaAsp 305
 DB 134 GCGGACACAGAGGTTGGCTTCGCGGAT---GGGCTTACA--AGCCAGAG-TGCTCC 186
 QY 305 roGlySerGlyGlnlleGlnLeuTrpGlnPheLeuLeuGlnLeuSerAspSerSera 325
 DB 187 CTGGCTCAAGGACGATCCAGCTGTGCACTTATCTGGAGCTGCTGGGAAGAGAGAGT 246
 QY 325 snSerAsnCysIleThrTrpGlnGlyThrAsnGlyGluPheLysMetThrAspProAspG 345
 DB 247 ACCAGGCGCTATTCCTGGCAGGGGAGGAGCTACGGGAATTCGATCAAGACCTGATG 306
 QY 345 luValAlaArgArgTrpGlyGluArgLysSerLysProAsnMetAsnTyraAspLysLeuS 365
 DB 307 AGTGAGCCCGCGCTGTGGGCGCTGCAAGTGCAAGCCCAAGATGAATGACACAAGCTGA 366
 QY 365 eraArgAlaLeuArgTyrrTyrrAspLysAsnIleMetThrLysValHisProGluS 385
 DB 367 GCCGGGCCCTGCTATTAATAACAAGCGCATTCACACAAGC----- 413
 QY 385 erSerMetTyrrLysTyrrProSerAspLeuProTyrrMetSerSerTyrrHisGlyLysArgT 405
 DB 414 -----AAGGGGAAACGCT 426
 QY 405 yrrAlaTyrrLysPheAspPheHisGlylleAlaGlnAlaLeuGlnProHis----- 421
 DB 427 TCACCTCAAGATTCATTCACAAACAGTGTGCTGATCAATTCATTCATTCATGATG 486

QY 422 -----AlaHisProGlnLysMetAsnPhenValAlaProHisProProAlaL 437
Db 487 GGTGGCTGGGGGTGACAGTCCCCAGAGT-----GCCCGCCGAG 525
QY 437 eubProValThrSerSerSerPhePheAlaAlaProAsnProTyrTyrPheSerProthrg 457
Db 526 TGCCTGCTGGGTGTAGCCACTCCGCTCCCTCC-----TCACAGCCCTCCG 573
QY 457 TylGlyLylTyrProAsnThrArgLeuProAlaAlaHisMetProSer---HisLeu 474
Db 574 A-GGTGTGTCTCCCAACCG-AGGACCCCGCTCACACACAGCTGCTTCATCTT 627
RESULT 11
US-08-343-443B-5
; Sequence 5, Application US/08343443B
; Patent No. 5968734
; GENERAL INFORMATION:
; APPLICANT: Aurias, Alain
; APPLICANT: Delattre, Olivier
; APPLICANT: Desmazes, Chantal
; APPLICANT: Melot, Thomas
; APPLICANT: Peter, Martine
; APPLICANT: Ploougaestel, Beatrice
; APPLICANT: Thomas, Gilles
; APPLICANT: Zucman, Jessica
; TITLE OF INVENTION: NUCLEIC ACID CORRESPONDING TO A GENE OF
; TITLE OF INVENTION: CHROMOSOME 22 INVOLVED IN RECURRENT CHROMOSOMAL
; TITLE OF INVENTION: TRANSLATIONS ASSOCIATED WITH THE DEVELOPMENT OF CANCEROUS
; TITLE OF INVENTION: TUMORS, AND NUCLEIC ACIDS OF FUSION RESULTING FROM SAID
; NUMBER OF SEQUENCES: 129
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Welser & Associates
; STREET: 230 South Fifteenth Street
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: AEDIT 1.0 DOS text editor
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/343,443B
; FILING DATE: 18-NOV-1994
; CLASSIFICATION: 514
; APPLICATION DATA:
; APPLICATION NUMBER: PCT/FR93/00494
; FILING DATE: 19-MAY-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 92/06123
; FILING DATE: 20-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Welser, Gerard J.
; REGISTRATION NUMBER: 19,763
; REFERENCE/DOCKET NUMBER: 989,6121P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-875-8383
; TELEFAX: 215-875-8394
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 328 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..327
; US-08-343-443B-5
Alignment Scores:

Pred. No.: 5.57e-22 Length: 328
Score: 301.50 Matches: 57
Percent Similarity: 79.31% Conservative: 12
Best Local Similarity: 65.52% Mismatches: 17
Query Match: 11.65% Indels: 1
DB: 2 Gaps: 1
US-09-902-772-4 (1-478) x US-08-343-443B-5 (1-328)
QY 252 TyrGlnGlnAlaArgArgSerAlaTrpThrSerHisSerHisProThrGlnSerLysAla 271
Db 67 TATGACACACTGCAGAT 126
QY 272 ThrGln---ProSerSerThrValProLysThrGlnAspGlnArgProGlnLeuAsp 290
Db 127 CTRCCCTTGAGAGGGGACAAAGACATGATGATGATGATGATGATGATGATGATGATGAT 186
QY 291 ProTyrGlnThrLeuGlyProThrSerSerArgLeuAlaAsnProGlySerGlyGlnIle 310
Db 187 CCGATATCAGATCCTGGGCGGACGACGACGATGCGCTGAGCCAACTGGAAACGGGGAGATC 246
QY 311 GlnLeuTrpGlnPheLeuLeuGlnLeuLeuSerAspSerSerAsnSerAsnSerGlyIleThr 330
Db 247 CAGCTGGGCAATTCCTCCGAGAGCTGCTCTCCGACAGCCGACAGCCGACAGCTATATCACC 306
QY 331 TrpGlnGlyThrAsnGlyGlu 337
Db 307 TGGAGAGGGACAAAGGGGAG 327
RESULT 12
US-08-875-944B-1
; Sequence 1, Application US/08875944B
; Patent No. 6096542
; GENERAL INFORMATION:
; APPLICANT: FUTINAGA, Kei
; APPLICANT: YOSHIDA, Koichi
; APPLICANT: HIGASHINO, Fumihito
; TITLE OF INVENTION: CANCER CONTROL
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.
; STREET: 624 Ninth Street N.W., Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/875,944B
; FILING DATE: 07-AUG-1997
; CLASSIFICATION: 514
; APPLICATION DATA:
; APPLICATION NUMBER: JP 07-020173
; FILING DATE: 08-FEB-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP96/00016
; FILING DATE: 09-JAN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWDY, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: FUTINAGA-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-5197
; TELEFAX: (202) 737-3528
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2064 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single

```

; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1386
; US-08-875-944B-1

Alignment Scores:
Pred. No.: 1.77e-20 Length: 2064
Score: 299.50 Matches: 136
Percent Similarity: 34.84% Conservative: 41
Best Local Similarity: 26.77% Mismatches: 141
Query Match: 11.57% Indels: 192
DB: 3 Gaps: 17

US-09-902-772-4 (1-478) x US-08-875-944B-1 (1-2064)
QY 45 SerIysMetSerProArgValProGlnGlnAspThrLeuSerGlnProAlaArgVal 64
DB 218 TCAGAAAGAGAGCCCGAGAGTCCCGCAG- 247
QY 65 ThrIleuMetGluCysAsnProAsnGlnValAsnGlySerArgAsnSerProAspAsp 84
DB 248 -----ACCCGGCCCTGTCCTGCAGCAGGAAGCCCGCCTCCCTAC- 289
QY 85 CysSerValAlaLysGlyGlyLysMetValSerSerSerAspAsnValGlyMetAsnTyr 104
DB 290 -----ACCATGGCGAGCAGTGC- 307
QY 105 GlySerTyrMetGluGluLysHisIleProProProAsnMetThrThrAsnGluArg-Arg 124
DB 308 -----TTTACTCTCAGTGCCTATGACCCCGCAGCAAAATCG 343
QY 124 GValIleValProAlaAspProThrLeuTyrSerThrAspHisValArgGlnTyrLeuG1 144
DB 344 CCATCAGATCCCTGC- 370
QY 144 UTPrAlaValLysGluTyrGlyLeuProAspValAspIleLeuLeuPheGlnAsnIleAs 164
DB 371 TTGGACAGTGCCTCAGAGCCCTTTCCTCCCGCAGAGCAAGGAAATTTCTCGAGATC- 428
QY 164 pGlyLysGluLeuLysMetThrLysAspAspPheGlnArgLeuThrProSerTyrAs 184
DB 428 ----- 428
QY 184 nAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGluThrProLeuProHisLeuTh 204
DB 429 -----CTCTGGCAGCTCCAGCCCGCAGCTCGCATGGGTA- 464
QY 204 rSerAspAspValAspLysAlaLeuGlnAsnSerProArg- 217
DB 465 -----CTCTGGGAGACATAGTCCGCTTCC 490
QY 218 -----LeuMetHisAlaArgAsnThrGlyGlyAlaThrPheI1 230
DB 491 AGCAGCCCTGGGACATTTGCCACTCTCCACATCTCAGGAGGAGG- 536
QY 230 ePheProAsnThrSerValTyrProGlnAlaThrGlnArgGlnThrArgProAspLe 250
DB 537 -----CCGGGAACCCCTCCAGCCCGCCTACCAACACAGCTGTCCGAGCCCGCCACCT 592
QY 250 u-----Pro-TyrGluGlnAlaArgArgS 258
DB 593 ATCCCGACAGAGCTTAAAGCAGATACATCATGATCCCTGTATAGACAGCGGCGAGC 652
QY 258 eAlaI-TyrThr-----Ser 262
DB 653 CAGCCGTGGACAGGCTTAAAGGAGGATAGGACACAGTACCCAGGGGCGGTGTGATCA 712
QY 263 HisSerHisProThrGlnSerLysAlaThrGlnProSerSerSerThrValProLysThr 282
DB 713 AACAGGAACAGAGGAGCTTGCCTACAGACTCAGATGTCACCGGGTGGCGCATCAATGTAAC 772

```

```

QY 283 GluAspGlnArgProGlnLeuAspProTyrGlnIleLeuGlyPro----- 297
DB 773 TCACACAGAGAGGCTTCTGTGGCCCTCTCCAGGTACAGGGGCGCATGGGCTATGGCTATG 832
QY 297 ----- 297
DB 833 AGAAACCTCTGCCAGCATTTCCACAGATGATGTCTGCTTCTCCCTGAGAAATTTGAGAG 892
QY 298 ThrSerSerArgLeuAlaAsnProGly-----SerGly 308
DB 893 ACATCAAGACAGAGAGGGGTGCGT-GCATTTGCAGAGGGGCGCCCTTACCAGCGCGGGGT 951
QY 309 GlnIleGlnLeuTyrPheGlnPheLeuLeuLeuLeuSerAspSerSerAsnSerAsnGly 328
DB 952 GCCCTGCAGCTGTGCATTTCTGTGCTGTGATGACCCAAATGCCCATTTC 1011
QY 329 IleThrTrpGluGlyThrAsnGlyLysPheLysMetThrAspProAspGluValAlaArg 348
DB 1012 ATTGCTTGAGAGCGGCGCGGGAATGAGATTCAAGCTCATTTGAGCTGAGAGAGGTGCGCAGG 1071
QY 349 ArgTrpGlyGluArgLysSerLysProAsnMetAsnTyrAspLysLeuSerArgAlaLeu 368
DB 1072 CTCTGGGCGATCCAGAGAACCGGCGACCATGATTAACAGCAAGCTGAGCGGCTGCTC 1131
QY 369 ArgTyrTyrTyrAspLysAsnIleMetThrLysValHisProProGluSerSerMetTyr 388
DB 1132 CGATACATATTTATGAAAGGCATCATCGAGAAGGTG--GCTGGTGAGCGTTAGCTGTAC 1188
QY 389 LysTyr-----ProSerAspLeuProTyrMetSerSerTyrHisGlyLysArgTyr 405
DB 1189 AAGTTTGTGTGTAGAGCCCGAGAGCCCTTCTTGTGGCTTCCCGGACAAATCAGCGTCA 1248
QY 406 AlaTyrLysPheAspPheHis-GlyIleAlaGlnAlaLeuGlnProHisAlaHisProG1 425
DB 1249 GCTCTCAAGCTGAGTTTGACCGGCTGTA-----GTGAG 1284
QY 425 nLysMetAsnPheValAlaProHisProProAlaLeuProValThrSerSerPhePh 445
DB 1285 GAGGACACAGTCCCTTGTCTCCACTTGATGAGAGCCCGCTCAGAGAGTGGCT 1344
QY 445 eAlaAlaProAsnProTyrTrpAsnSerProThrGlyLysIleTyr-----ProAsnTh 463
DB 1345 GGCCCGCCCGCAG-CCATTT-----GGCCCAAGGGGTGCTACTGTACTAGCCCGCAGC 1397
QY 463 rArgLeuProAlaAla 468
DB 1398 GGCTGTTCCTCCCTGCC 1413

RESULT 13
US-09-116-049-3
; Sequence 3, Application US/09116049A
; Patent No. 6248351
; GENERAL INFORMATION:
; APPLICANT: Hung, Men-Chie
; TITLE OR INVENTION: HUMAN PEA3 IS A TUMOR SUPPRESSOR FOR CANCER CELLS
; FILE REFERENCE: US/09-116-049A
; CURRENT APPLICATION NUMBER: US/09/116, 049A
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 2064
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-116-049-3

Alignment Scores:
Pred. No.: 1.77e-20 Length: 2064
Score: 299.50 Matches: 136
Percent Similarity: 34.84% Conservative: 41
Best Local Similarity: 26.77% Mismatches: 141
Query Match: 11.57% Indels: 192
DB: 4 Gaps: 17

```



```

: LENGTH: 1905 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: IMMEDIATE SOURCE:
: LIBRARY: PROSUT12
: CLONE: 1813005
US-09-055-113-2

Alignment Scores:
Pred. No.: 2,48e-20
Score: 297.50
Percent Similarity: 34.86%
Best Local Similarity: 25.47%
Query Match: 11.50%
Gaps: 4
Indels: 159
Matches: 122
Conservative: 45
Mismatch: 153
Indels: 159
Gaps: 20

US-09-902-772-4 (1-478) x US-09-055-113-2 (1-1905)

QY 41 TyrGlyGlnThrSerLysMetSerProArgValProGlnGlnAspTrpLeuSerGlnPro 60
Db 628 TACCTTGAGGAGACAGCTGGGACAGCCCTGGGGCCAGCAGTCGGGAGAGAGCA 687
QY 61 ProAlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsn 80
Db 688 CCT-----GAG 693
QY 81 SerProAspAspCysSerValAlaLysGlyLysMetValSerSerSerAspAsnVal 100
Db 694 GAGCGTGAGCAGTGGCCGGTCATTGACAGCCAGCCCGGAGCGGAGCTGGACTTGGTGT 753
QY 101 GlyMetAsnTrpGlySerTrpMetGluGluLysHisIle-----ProProProAsnMet 118
Db 754 CCC-----GGGGGGCTGACCTTGGAGAGGACCTGCTGAGACAGAGTCCAGCATG 804
QY 119 ThrThrAsnGlu-----ArgArgValIleValProAla 129
Db 805 GTGGTGGGCGAAGTCTCTAAGACATCGAAGCGGCTTCAGACCTGCTCAACCTCACCCCA 864
QY 130 AspProThrLeuTrpSerThrAspHisValArgGlnTrpLeuGluTrpAlaValLysGlu 149
Db 865 GATCCCAATGAGACTGGAGCCCAAGCAATGTGCAGAGTGGCTGCTGTGACAGAGACCA 924
QY 150 TyrGlyLeuProAspValAspIleLeuLeuPheGlnAsnIleAspGlyLysGluLeuCys 169
Db 925 TACCGGCTGCCCCCATGGGC---AAGGCTTCCAGGAGCTGGCGGCAAGGAGCTGTGC 981
QY 170 LysMetThrLysAspAspPheGlnArgLeuThrProSerTrpAsnAlaAspIleLeuLeu 189
Db 982 GCCATGTGGAGAGACACTTCCGACGCGCTGCCCC---CTGGGTGGGAGTGTCTGCAC 1038
QY 190 SerHisLeuHisTrpLeuArgGluThrProLeuProHisLeuThrSerAspAspValAsp 209
Db 1039 GCCCACCTG-----GACATCTGG 1056
QY 210 LysAlaLeuGlnAsnSerProArgLeuMetHisAlaArgAsnTrpGlyValAlaThrPhe 229
Db 1057 AAG-----TCAGCGGCGCTGATGAAGACCGAGCTTCACTCGGCGCATTCAC 1104
QY 230 IlePheProAsnThrSerValTrpProGlnAlaThrGlnAlaGlyIleThrThrArgProAsp 249
Db 1105 TACTGTGCTGCTGACACAGT----- 1122
QY 250 LeuProTrpGluGlnAlaArgSerAlaTrpThrSerHisSerHisProThrGlnSer 269
Db 1123 -----GAGGAGAGCTGG 1134
QY 270 LysAlaThrGlnProSerSerThrValProLysThrGluAspGlnArgProGlnLeu 289
Db 1134 ----- 1134
QY 290 AspProTrpGlnIleLeuGlyProThrSerSerArgLeuAlaAsnProGlySerGlyGln 309
Db ----- 309

```

```

Db 1135 -----ACGACAGCAGAGTGGAGTCAATCATCTCCGGCAG 1170
QY 310 ---IleGlnLeuTrpGlnPheLeuLeuGlu---LeuLeuSerAspSerSerAsnSerAsn 327
Db 1171 CCCATCCACACTGTGGCAGATTCTCAAGGAGTGTACTACCAAGCCCAAGCATATGAGCCGC 1230
QY 328 CysIleThrTrp---GluGlyThrAsnGlyGluPheLysMetThrAspProAspGluVal 346
Db 1231 TTCATTAGGTGGCTCCACAGAGGAGGACATCTTCAAAATGTAGAGACTCACCCAGGTG 1290
QY 347 AlaArgArgTrpGlyGluArgLysSerLysProAsnMetAsnTrpAspLysLeuSerArg 366
Db 1291 GCCCGGCTGTGGGAGATCGGCAAGAACCGTCCCGCATGAATACAGCAAGCTGAGCCGC 1350
QY 367 AlaLeuArgTrpTrpTrpAspLysAsnIleMetThrLys----- 379
Db 1351 TCCATCCGCGCAGTATTACAGAGGAGGATCATCCGGAAGCCAGACATCTCCAGGCGCTC 1410
QY 380 -----ValHisPro-----ProGlnSerSerMetTrp 388
Db 1411 GTCTACAGATTCTGTGTCACCCCATCTGAGTGCCTGGCCGAGGCGCTGAACCCGCTCAG 1470
QY 389 LysTrpProSerAspLeuProTrpTrpMetSerSerTrpHisGlyLysArgTrpAlaTrpLys 408
Db 1471 GGGCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1512
QY 409 PheAspPheHisGlyIleAlaGlnAlaLeuGlnProHisAlaHisProGlnLysMetAsn 428
Db 1513 TGGGGGAAAGAGGAGCATGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 1563
QY 429 PheValAlaProHisProProAlaLeuPro----- 438
Db 1564 ---GTACAGAGAGGAGGCAACCACTGCCAGGGGAGATAGGGTCTGTGGGCTTGGGG 1620
QY 439 ---ValThrSerSerSerPhePheAlaAlaProAsnProTrpTrpAsnSerProThrGly 457
Db 1621 ACCCTGGGCGAGGGTGTCTCTCTCTGCTGAGGCCAGC-----TGCCTCCCTGAGAGA 1671
QY 458 -----GlyIleTrpProAsnThrArgLeuProAlaAlaHisMetProSer 472
Db 1672 CAGAGGAGAGACAGGGCTGTCTCTCCCAACACTGCTGACCCGACCATTTCCAGAGC 1728

RESULT 15
US-08-780-835B-1
: Sequence 1, Application US/08780835B
: Patent No. 5922688
: GENERAL INFORMATION:
: APPLICANT: Hung, Mien-Chie
: APPLICANT: King, Xiangming
: TITLE OF INVENTION: PEA3 is a Tumor Suppressor
: NUMBER OF SEQUENCES: 9
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: ARNOLD, WHITE AND DURKEE
: STREET: P.O. Box 4433
: CITY: Houston
: STATE: Texas
: COUNTRY: USA
: ZIP: 77210-4433
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/780,835B
: FILING DATE: 10-JAN-1997
: CLASSIFICATION: 514
: ATTORNEY/AGENT INFORMATION:
: NAME: Wilson, Mark B.
: REGISTRATION NUMBER: 37,259
: REFERENCE/DOCKET NUMBER: UMSC500
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (512) 418-3000

```

```

? TELEFAX: (512) 474-7577
? INFORMATION FOR SEQ ID NO: 14
? SEQUENCE CHARACTERISTICS:
? LENGTH: 2410 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
;
;
; US-08-780-835B-1

```

Alignment Scores:	
pred. No.:	2,42e-18
Score:	279.50
Percent Similarity:	35.22%
Best local Similarity:	25.53%
Query Match:	10.80%
DB:	2
Length:	2410
Matches:	108
Conservative:	41
Mismatches:	115
Indels:	160
Gaps:	15

US-09-902-772-4 (1-478) x US-08-780-835B-1 (1-2410)

QY	45	SerIysMetSerProAArgValProGInGlnAspTrpLeuSerGInProProAlaIArgVal	64
Db	634	TCAAGAAAGAAACCCCAAGTCCC-----GCACAGACCCGCCCTGTCT	678
QY	65	ThrIleuYMetGluCysAsnProAsnGlnValAsnGlySerAArgAsnSerProAsp---	83
Db	679	GCACAGGAAGCCACACCTCCCTACACACATGGAGAGAGAGTGCTCCCTTACTCCAGACAA	738
QY	84	-----AspCysSerValAlaIlys	89
Db	739	TCCGCATCAAGTCCCCCGCTCCCGGTGCCCTGGACAGTGGCCCCCTTGACACCCCTTTTCCA	798
QY	90	GlyGlyIysMetValSerSerSerAspAsnValGlyMetAsnTrpGlySerTrpMetGlu	109
Db	799	GGGCAAGAACACACAGACAGCTCC-----TGAGAG	828
QY	110	GluIysHisIleProProAsnMetThrThrasnGluArgArgValIleValProAla	129
Db	829	CCTCACACTCTTCCCACTCCACCTGGCCACGGGTACTTGGTAGACACAGCTCCGTCT	888
QY	130	AspProThrLeuTrpSerThrAspHisValArgGlnTrpLeuGluTrpAlaValLysGlu	149
Db	889	TCCAGACGCCCGTGACA-----	906
QY	150	TyrGlyLeuProAspValAspIleLeuLeuPheGlnAsnIleAspGlyLysGluLeuCys	169
Db	907	-----TGT	909
QY	170	LysMetThrLysAspAspPheGlnArgLeuThrProSerTrpAsnIleAspIleLeuLeu	189
Db	910	-----GCCACTCTTCA-----	921
QY	190	SerHisLeuHisTrpLeuArgGluThrProLeuProHisIleuThrSerAspAspValAsp	209
Db	922	-----CATCTCCACAGAG--GGGGCGGGAACTCTCCAGCGCCCTATCACC	969
QY	210	LysAlaLeuGlnAsnSerProArgLeuMetHisAlaArgAsnThrGlyAlaThrPhe	229
Db	970	ACCAACTGTGTGGAGCCCTGGCCACACCTTACCCCAAGAACTTCAAGACAGAGATACCATTG	102
QY	230	IlePheProAsnTrpSerValTyrProGluAlaThrGlnArgIleThrThrArgProAsp	249
Db	1030	-----ACCCCTGTACAAACGCTGGCCACG-----	105
QY	250	LeuProTrpGlnGlnAlaArgSerAla-----Trp	260
Db	1057	-----CCGCTT---CAAGCCAGAGTGGGGTCACTGTGGCAGACAGTCCACAGGGCGGGGGTGG	1111
QY	261	ThrSerHisSerHisProThrGlnSerLysAlaThrGlnProSerSerSerThrValPro	280
Db	1111	TGATCAAAACAGAGCGCACAGACTTGGCGCTAGACATCAAGATGTCCCTGGATGTGCATCA	117
QY	281	LysThrGluAspGlnArgProGlnLeuAspProTyrGlnIle-----	294

Db	1171	TGTACGTCAACCCAGAGGGCTTCTGTGAAACCCCTCCAGGTGATGGAGTGTATNG	1230
Qy	295	-----Leu	295
Db	1231	GCTATGAAAAATCCCTTCGACCATTCACAGATGATGTCTGCATTGCTCCATAAAATTTG	1290
Qy	296	GLYProHisSerSerArg-----LeuAlaAsnProGly-----	306
Db	1291	AAGGAGACATTCACAGACAGAGGATTTGGAGCTTTCCGGAG-GGGCCACCCACACAGGC	1349
Qy	307	SerGlyGlnIleGlnIleuTrpGlnPheLeuLeuGlnLeuLeuSerAspSerSerAsnSer	326
Db	1350	CGGGGTCCCTTACACAGTGGCGAGTTTCGTGGTGGCCCTCTCGATGACCCACCAACAAATCT	1409
Qy	327	AsnGlyIleHrTrpGlnGluGlyThrAsnGlyGluPheLeuMetThrAspProAspGluVal	346
Db	1410	CATTTCATTGCTTGGACAGCGCCGGGGAAGAAGATTTAAACTAATTGMACTGAAGAGTT	1469
Qy	347	AlaArgArgTrpGlyGluArgGlySerLysProAsnMetAsnTyrAspLysLeuSerArg	366
Db	1470	GCCGCGCTCGGGGTATCCAGAGAACCGGCCCATTAATTAATGACACGCTGAGCGCG	1529
Qy	367	AlaLeuArgTyrTyrTyrAspLysAsnIleMetThrLysValHisProProGluSerSer	386
Db	1530	TCCGTGCGATCACTATTATGTGAAAGGCATCATGACAGAGTG-----	1571
Qy	387	MetTyrLysTyrProSerAspLeuProTyrMetSerSerTyrHisGlyLysArgTyrAla	406
Db	1572	-----GCTGGGGAACGCTACGTG	1589
Qy	407	TyrTyrPhe	409
Db	1590	TACAAATTT	1598

Search completed: November 9, 2002, 19:49:16
Job time : 75.0549 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 9, 2002, 18:55:38 ; Search time 68.9473 Seconds
(without alignments)
2459.669 Million cells updates/sec

Title: US-09-902-772-4
Perfect score: 2588
Sequence: 1 MASTIKREALSVSEDSLEF.....ITPNRLPAHPMSHLCTTY 478

Scoring table: BIOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 320260 segs, 177392727 residues

Total number of hits satisfying chosen parameters: 640520

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+ p2n.model -DEV=x1p
-O/cgn2_1/USPTO.spool/US09902772/runat.07112002.134035.9131/app.fasta.1.1294
-DB=Published.Applications.NA -QFMT=fastap -SUFFIX=rmpb MINMATCH=0.1
-LOOPCUT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=BIOSUM62
-TRANS=human40.ccl -LIST=45 -DOCCALIGN=200 -THR.SCORE=pct -THR.MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAJLEN=2000000000 -USER=US09902772.6CGN.1.1.46.07112002.134035.9131
-NCPU=6 -ICPU=3 -NO.XLPHY -NO.MMAP -LARGEORDERY -NEG.SCORES=0 -WAIT -LONGLOG
-DRV.TIMEOUT=120 -MARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published.Applications.NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
c 1	801	31.0	567 10	US-09-864-761-18410 Sequence 18410, A
c 2	716	27.7	473 10	US-09-864-761-20472 Sequence 20472, A
c 3	565	21.8	549 10	US-09-923-779-54 Sequence 54, Appl
c 4	503	19.4	420 10	US-09-864-761-3705 Sequence 3705, Ap

c 5	503	19.4	454 10	US-09-864-761-1653 Sequence 1653, Ap
c 6	481	18.6	1752 10	US-09-850-799-1 Sequence 1, Appl1
c 7	444	17.2	2168 10	US-09-920-300A-1716 Sequence 1716, Ap
c 8	444	17.2	2168 12	US-10-033-528-1716 Sequence 1716, Ap
c 9	444	17.2	2168 10	US-09-920-300A-1693 Sequence 1693, Ap
c 10	444	17.2	2268 12	US-10-033-528-1693 Sequence 1693, Ap
c 11	430.5	16.6	1884 10	US-09-925-300-420 Sequence 420, App
c 12	382	14.8	348 10	US-09-864-761-29364 Sequence 29364, A
c 13	382	14.8	477 10	US-09-864-761-5035 Sequence 5035, Ap
c 14	382	14.8	519 10	US-09-864-761-12798 Sequence 12798, A
c 15	374	14.5	553 10	US-09-864-761-7427 Sequence 7427, Ap
c 16	371	14.3	225 10	US-09-864-761-21761 Sequence 21761, A
c 17	305	11.8	3178 9	US-10-108-605-124 Sequence 124, App
c 18	305	11.8	3178 9	US-10-108-605-128 Sequence 128, App
c 19	299.5	11.6	2333 10	US-09-920-300A-1788 Sequence 1788, Ap
c 20	299.5	11.6	2333 10	US-09-880-107-3316 Sequence 3316, Ap
c 21	299.5	11.6	2333 12	US-10-033-528-1788 Sequence 1788, Ap
c 22	297.5	11.5	1894 10	US-09-126-945B-1 Sequence 1, Appl1
c 23	297.5	11.5	1905 10	US-09-866-356-2 Sequence 2, Appl1
c 24	280.5	10.8	2952 10	US-09-925-297-257 Sequence 257, App
c 25	279	10.8	1894 10	US-09-864-761-23294 Sequence 23294, A
c 26	278.5	10.8	165 10	US-09-841-963A-1 Sequence 1, Appl1
c 27	244.5	9.4	1426 10	US-09-925-297-309 Sequence 309, App
c 28	244	9.4	1429 10	US-09-764-864-320 Sequence 320, App
c 29	236	9.1	1704 10	US-09-841-963A-3 Sequence 3, Appl1
c 30	233	9.0	4190 10	US-09-962-832-111 Sequence 111, App
c 31	217	8.4	1996 10	US-09-925-301-207 Sequence 207, App
c 32	215	8.3	1915 10	US-09-964-824A-101 Sequence 101, App
c 33	215	8.3	1915 10	US-09-964-824A-563 Sequence 563, App
c 34	215	8.3	1915 10	US-09-880-107-3420 Sequence 3420, Ap
c 35	215	8.3	1915 10	US-09-967-768A-192 Sequence 192, App
c 36	215	8.3	1917 10	US-09-922-217-1105 Sequence 1105, App
c 37	213.5	8.2	852 10	US-09-759-143-44 Sequence 44, Appl
c 38	213.5	8.2	852 10	US-09-759-143-44 Sequence 44, Appl
c 39	213.5	8.2	852 10	US-09-030-606-44 Sequence 44, Appl
c 40	213.5	8.2	852 10	US-09-822-827-44 Sequence 44, Appl
c 41	213.5	8.2	852 10	US-09-115-453-44 Sequence 44, Appl
c 42	194	7.5	2642 10	US-09-920-300A-1680 Sequence 1680, Ap
c 43	194	7.5	2642 12	US-10-033-528-1680 Sequence 1680, Ap
c 44	185.5	7.2	488 10	US-09-777-564-468 Sequence 468, App
c 45	184.5	7.1	507 10	US-09-777-564-489 Sequence 489, App

ALIGNMENTS

RESULT 1
US-09-864-761-18410/c
Sequence 18410, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aeomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263,6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30

```

; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 18410
; LENGTH: 567
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AF000163.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN HELI00, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN HEPA, SIGNAL = 0.92
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.95
; OTHER INFORMATION: SWISSPROT HIT: P1308, EVALU0 9.00e-99
; OTHER INFORMATION: EST_HUMAN HIT: R87572.1, EVALU0 0.00e+00
; OTHER INFORMATION: NT HIT: M17254.1, EVALU0 0.00e+00
US-09-864-761-18410

Alignment Scores:
Pred. No.: 1.34e-70 Length: 567
Score: 801.00 Matches: 151
Percent Similarity: 78.17% Conservative: 3
Best Local Similarity: 76.65% Mismatches: 3
Query Match: 30.95% Indels: 40
DB: 10 Gaps: 2

US-09-902-772-4 (1-478) x US-09-864-761-18410 (1-567)
QY 302 LeuAlaAsnProGlySerGlyInleGlnLeuTTPGlnPheLeuLeuGlnLeuLeuSer 321
||| : : : |||
DB 534 GCACCTCATTTAGGAGTGGCCAGATCCAGCTTGGCACTTCCTCGAGCTCCCTGTG 475
||| : : : |||
QY 322 AspSerSerAsnSerAsnGlyIleThrTrpGluGlyThrAsnGlyGlnPheLeuMetThr 341
||| : : : |||
DB 474 GACACTCCAACTCCAGCTGATCAGCTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 415
||| : : : |||
QY 342 AspProAspGluValAlaAlaArgTrpGlyGluArgGlySerGlyProAsnMetAsnTrp 361
||| : : : |||
DB 414 GATCCCGAGAGAGTGGCCCGGCGCTGGGGAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAG 355
||| : : : |||
QY 362 AspLysLeuSerArgAlaLeuArgTrpTrpTrpAspLysAsnIleMetThrLysValHis 381
||| : : : |||
DB 354 GATTAAGCTCAGCGCGCCCTCCGTTACTACTATGACAAAGACATCATGACCAAGGTC--- 298
||| : : : |||
QY 382 ProProGluSerSerMetLysTrpTrpProSerAspLeuProTyrMetSerSerTrpHis 401
||| : : : |||

```

```

DB 297 -----CAT 295
QY 402 GlyLysArgTyrAlaLysLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnPro---- 420
||| : : : |||
DB 294 GGGAAACGGCTAGCGCTACAGAGTTGAGCTTCCAGCGGATGCGCCAGGCGCTCCAGCGCCAC 235
||| : : : |||
QY 421 -----His 421
DB 234 CCCCCGAGTCATCTCTGTACAAGTACCCCTGAGACCTCCCGTACATGGGCTCTCTATCAC 175
||| : : : |||
QY 422 AlaHisProGlnLysMetAsnPheValAlaProHisProProAlaLeuProValThrSer 441
||| : : : |||
DB 174 GCCACCCACAGAGATGAACTTTGTGGCCCCCAGCCCTCCAGCCCTCCGATCATCT 115
||| : : : |||
QY 442 SerSerPhePheAlaAlaProAsnProTyrTrpAsnSerProThrGlyGlyIleTrpPro 461
||| : : : |||
DB 114 TCCAGTTTTCCTGCTGCCCAAAACCATATCTGGAATTCACCAAGTGGGGTATATACCCC 55
||| : : : |||
QY 462 AsnThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTrp 478
||| : : : |||
DB 54 AACACTAGGCTCCGCCACAGCATATGCTCTCATCTGGCAGCTACTAC 4

RESULT 2
US-09-864-761-20472/C
; Sequence 20472, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Aecolica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117

```

```
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 20472
; LENGTH: 473
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AP000021.2
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.97
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1
; OTHER INFORMATION: EST_HUMAN HIT: R87572.1, EVALU0 0.00e+00
; OTHER INFORMATION: SWISSPROT HIT: P11308, EVALU0 6.00e-88
; OTHER INFORMATION: NT HIT: M17254.1, EVALU0 0.00e+00
US-09-864-761-20472

Alignment Scores:
Pred. No.: 2,56e-62 Length: 473
Score: 716.00 Matches: 136
Percent Similarity: 77.40% Conservative: 1
Best Local Similarity: 76.84% Mismatches: 0
Query Match: 27.67% Indels: 40
DB: 10 Gaps: 2

US-09-902-772-4 (1-478) x US-09-864-761-20472 (1-473)
QY 306 GYSGRLYGLNLEGINLEUTPGINPHELEULEUGLEULEUSERASPSERASN 325
DB 473 GGAGGTGGCCAGAGCCAGCTTGGCAGTTCCTCGAGCTCCGTGGAGCAGCTCCAC 414
QY 326 SERASNCYSILIEHTRTPGLUGLYTHRASNGLYGLUPHELYSMETTHRASPPOASPLU 345
DB 413 TCCAGCGCATCACTCGGAGAGGACCAACAGCGGAGTTCAGATGACGGATCCGACGAG 354
QY 346 VALAALARGATRGTPGLUGLYUARGLYSERLYSPROASNMETASNTRYASPLYLEUSER 365
DB 353 GTGGCCCGCGCTGGGAGAGCGGAGCGGAAGCAACCCCAATGAACTAGATTAAGCTCAGC 294
QY 366 ARGALAEUARGTYRTRYRYSRPLYSANLLEKETHRYSVALHISPROGLOUSER 385
DB 293 CGGCGCTCGGTACTATGACAAAGACATCATGACCAAGGTC----- 249
QY 386 SEMETLYLSTYRPROSERASPLEUPROTYRMEISERTYRHSGLYLSARGTYR 405
DB 248 -----CATGGGAAGCGCTAC 234
QY 406 ALATYRLYSPHEASPHENHISGLYILEAGLHIALALEUINPRO----- 420
DB 233 GCTACAAAGTTGACCTCCACGGGATGCGCCAGGCCCTCCAGGCCACCCCGGAGTCA 174
QY 421 -----HISALAHISPROGLN 425
DB 173 TCTCTGTACAAGTACCCCTCAGACCTCCGTACATGGGCTCTATCAGCCACCCACAG 114
QY 426 LYSMETASNPHYALALAPROHISPROPROALALEUPROVALHRSERSETSERPHEPHE 445
DB 113 AAGATGAACCTTGTTGGGGCCCCCAGCCCTCCAGCCCTCCCGTACATTTTCAAGTTT 54
QY 446 ALALAPROASNPROTYRTPASNSETPROTHGLYGLYLETYRPROASN 462
DB 53 GCTGCCCAAAACCATCTGGAATTCACCAACTGGGGGTATATACCCCAAC 3
;
RESULT 3
US-09-923-779-54
; Sequence 54, Application US/09923779
; Patent No. US2002007672A1
; GENERAL INFORMATION:
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiaochun
```

```
APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF PANCREATIC CANCER
; FILE REFERENCE: 210121.553
; CURRENT APPLICATION NUMBER: US/09/923,779
; CURRENT FILING DATE: 2001-08-06
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 54
; LENGTH: 549
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 356, 363, 417, 433, 442, 461, 463, 464, 469, 479, 485, 489,
; LOCATION: 537, 545
; OTHER INFORMATION: n = A,T,C or G
US-09-923-779-54

Alignment Scores:
Pred. No.: 2,63e-47 Length: 549
Score: 565.00 Matches: 110
Percent Similarity: 71.15% Conservative: 1
Best Local Similarity: 70.51% Mismatches: 4
Query Match: 21.83% Indels: 41
DB: 10 Gaps: 2

US-09-902-772-4 (1-478) x US-09-923-779-54 (1-549)
QY 344 ASPLUVALAALARGATRGTPGLUGLYUARGLYSERLYSPROASNMETASNTRYASPLYS 363
DB 2 GACAGAGTGGCCCGCCCTGGGAGAGCGGAAGCAACCAACATGAACTAGATTAAGTAAAG 61
QY 364 LEUSERARGALAEUARGTYRTRYRYSRPLYSANLLEKETHRYSVALHISPROPRO 383
DB 62 CTCAGCGCGCCCTCCCTTACTACTATGACAAACATCATGACCAAGTC----- 112
QY 384 GLUSERSETMYLSTYRPROSERASPLEUPROTYRMEISERTYRHSGLYLS 403
DB 113 -----CATGGGAAG 121
QY 404 ARGTYRVALATYRLYSPHEASPHENHISGLYILEAGLHIALALEUINPRO----- 420
DB 122 CGTACGCGCTTACAAAGTTGACCTTCACGGATGCGCCAGGCCCTCCAGCCACCCCG 181
QY 421 -----HISALAHIS 423
DB 182 GAGTCATCTCTGTAACAAGTACCCCTCAGACCTCCCGTACATGGGCTCTATCAGCCAC 241
QY 424 PROGLNLYSMETASNPHYALALAPROHISPROPROALALEUPROVALHRSERSETSER 443
DB 242 CCACAGAAAGATGAACCTTGTTGGGGCCCCCAGCCCTCCAGCCCTCCCGTACATTTTCA 301
QY 444 PHEPHEALALAPROASNPROTYRTPASNSETPROTHGLYGLYLETYRPROASNTHR 463
DB 302 TTTTTCCTGCCCCCAACCATATCTGGAATTCACCAACTGGGGGTATATACCCCTACT 361
QY 464 ARGLEUPROALAHISMETPROSERHISLEU-GLYTHR-TRYTYR 478
DB 362 ANGCTCCCAACAGCATATGCTTTTCACTTGGGGCACTTACTAC 407
;
RESULT 4
US-09-864-761-3705/C
; Sequence 3705, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David R.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Neomica-X-1
```

```

1 CURRENT APPLICATION NUMBER: US/09/864,761
2 CURRENT FILING DATE: 2001-05-23
3 PRIOR APPLICATION NUMBER: US 60/180,312
4 PRIOR FILING DATE: 2000-02-04
5 PRIOR APPLICATION NUMBER: US 60/207,456
6 PRIOR FILING DATE: 2000-05-26
7 PRIOR APPLICATION NUMBER: US 09/632,366
8 PRIOR FILING DATE: 2000-08-03
9 PRIOR APPLICATION NUMBER: GB 24263,6
10 PRIOR FILING DATE: 2000-10-04
11 PRIOR APPLICATION NUMBER: US 60/236,359
12 PRIOR FILING DATE: 2000-09-27
13 PRIOR APPLICATION NUMBER: PCT/US01/00666
14 PRIOR FILING DATE: 2001-01-30
15 PRIOR APPLICATION NUMBER: PCT/US01/00667
16 PRIOR FILING DATE: 2001-01-30
17 PRIOR APPLICATION NUMBER: PCT/US01/00664
18 PRIOR FILING DATE: 2001-01-30
19 PRIOR APPLICATION NUMBER: PCT/US01/00669
20 PRIOR FILING DATE: 2001-01-30
21 PRIOR APPLICATION NUMBER: PCT/US01/00665
22 PRIOR FILING DATE: 2001-01-30
23 PRIOR APPLICATION NUMBER: PCT/US01/00668
24 PRIOR FILING DATE: 2001-01-30
25 PRIOR APPLICATION NUMBER: PCT/US01/00663
26 PRIOR FILING DATE: 2001-01-30
27 PRIOR APPLICATION NUMBER: PCT/US01/00662
28 PRIOR FILING DATE: 2001-01-30
29 PRIOR APPLICATION NUMBER: PCT/US01/00661
30 PRIOR FILING DATE: 2001-01-30
31 PRIOR APPLICATION NUMBER: PCT/US01/00670
32 PRIOR FILING DATE: 2001-01-30
33 PRIOR APPLICATION NUMBER: US 60/234,687
34 PRIOR FILING DATE: 2000-09-21
35 PRIOR APPLICATION NUMBER: US 09/608,408
36 PRIOR FILING DATE: 2000-06-30
37 PRIOR APPLICATION NUMBER: US 09/774,203
38 PRIOR FILING DATE: 2001-01-29
39 NUMBER OF SEQ ID NOS: 49117
40 SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
41 SEQ ID NO 3705
42 LENGTH: 420
43 TYPE: DNA
44 ORGANISM: Homo sapiens
45 FEATURE:
46 OTHER INFORMATION: MAP TO AP000021.2
47 OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.2
48 OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1
49 OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1
50 OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.97
51 OTHER INFORMATION: EXPRESSED IN BL100, SIGNAL = 1.5
52 OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1
53 US-09-864-761-3705
54 Alignment Scores:
55 Pred. No.: 2,36e-41 Length: 420
56 Percent Similarity: 503.00 Matches: 96
57 Best Local Similarity: 69.57% Mismatches: 1
58 Query Match: 19.44% Indels: 40
59 Gaps: 2
60 US-09-902-772-4 (1-478) x US-09-864-761-3705 (1-420)
61 Oy 361 TyrasplylseuserAargalaleuAryTyTyTyTyrasplysaniIemethrLysval 3800
62 Db 418 TACCATAGCTCACCGCGGCGCTCCGTACTACTATGACAAACATCATGACCAAGTTC 359
63 Oy 381 HisProProgluserSerMetTyLysTyTyProseraspLeuproTyfMetSerSetTy 400
64 Db 359 ----- 359

```

```

QY      401 HisGlyLysArgTyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaIleuInPro 420
      |||
Db      358 CATGGGAAGCCGCTACGCTACAAAGTTCGACTTCCACGGGATGCCAGGCCCTCCAGGCC 299

QY      420 ----- 420

Db      298 CACCCCGGAGTCATCTGTACAGTACCCTTCAGACTCCCGTACATGGGCTCTAT 239

QY      421 HisAlaHisProGlnLysMetAsnPheValAlaProHisProProAlaLeuProValThr 440
      |||
Db      238 CACGCCACCCACAGAAATGAACCTTGTGGCGCCGCCACCTCCAGCCCTCCGTGACA 179

QY      441 SerSerSerPhePheAlaAlaProAsnProTyrTyrPAsnSerProThrGlyGlyIleTyr 460
      |||
Db      178 TCTTCACATTTTTTTTGTGCGCCCAAAACCATACGTGAAATTACCAACACTGGGGGRTATATAC 119

QY      461 ProAsnThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyr 478
      |||
Db      118 CCCAACTAGAGCTCCACGACCATATGCTTCTCATCTGGGCACTTACTAC 65

RESULT 5
US-09-864-761-1653/c
; Sequence 1653, Application US/09864761
; Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIORITY APPLICATION NUMBER: US 60/180,312
PRIORITY FILING DATE: 2000-02-04
PRIORITY APPLICATION NUMBER: US 60/207,456
PRIORITY FILING DATE: 2000-05-26
PRIORITY APPLICATION NUMBER: US 09/632,366
PRIORITY FILING DATE: 2000-08-03
PRIORITY APPLICATION NUMBER: GB 24263.6
PRIORITY FILING DATE: 2000-10-04
PRIORITY APPLICATION NUMBER: US 60/236,359
PRIORITY FILING DATE: 2000-09-27
PRIORITY APPLICATION NUMBER: PCT/US01/00666
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00667
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00664
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00669
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00665
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00668
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00663
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00670
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: US 60/234,687
PRIORITY FILING DATE: 2000-09-21
PRIORITY APPLICATION NUMBER: US 09/608,408
PRIORITY FILING DATE: 2000-06-30
PRIORITY APPLICATION NUMBER: US 09/774,203
PRIORITY FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1

```

```

; SEQ ID NO 1553
; LENGTH: 454
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AP000163.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.92
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.95
US-09-864-761-1653

Alignment Scores:
Pred. No.: 2.65e-41 Length: 454
Score: 503.00 Matches: 96
Percent Similarity: 70.29% Conservative: 1
Best Local Similarity: 69.57% Mismatches: 1
Query Match: 19.44% Indels: 40
DB: 10 Gaps: 2

US-09-902-772-4 (1-478) x US-09-864-761-1653 (1-454)
QY 361 TyrAspLysLeuSerArgAlaLeuAlaGlyTyrTyrAspLysAsnIleMetThrLysVal 380
Db 452 TACATAAGCTCAGCCGCGCTCCTGACTACTATGACAAAGACATATACCAAGGTC 393
QY 381 HisProGluSerSerMetTyrLysTyrProSerAspLeuProTyrMetSerSerTyr 400
Db 393 ----- 393
QY 401 HisGlyLysArgTyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnPro 420
Db 392 CATGGGAAGCGCTACGCTTACAGTTCGACCTCCAGGAGCGCCAGCGCTCCAGCCC 333
QY 420 ----- 420
Db 332 CACCCCGGAGTCATCTGTACAGTACCCCTCAGACCTCCGCTACAGTGGCTCTAT 273
QY 421 HisAlaHisProGlnLysMetAsnPheValAlaProHisProAlaLeuProValThr 440
Db 272 CAGCCCAACCCACAGAGATGAATCTTGTGGCGCCCAACCTCCAGCCCTCCCGTGACA 213
QY 441 SerSerSerPhePheAlaAlaProAsnProTyrTyrPheSerProThrGlyIleTyr 460
Db 212 TCTTCAGATTCTTGTGCTGCCCCAAACCCATATCGAATTCCACCAACTGGGGGTATATAC 153
QY 461 ProAsnThrArgLeuProAlaAlaHisMetProSerHisLeuGlyIleTyrTyr 478
Db 152 CCCAACACTAGGCTCCACCAACGCAATATGCTTCTCATCTGGGCACTACTATAC 99

RESULT 6
US-09-850-799-1
; Sequence 1, Application US/09850799
; Patent No. US20020090647A1
; GENERAL INFORMATION:
; APPLICANT: Deneris, Evan S.
; APPLICANT: Pyodero, Dmitry V.
; APPLICANT: Hendricks, Timothy J.
; TITLE OF INVENTION: Reagents and Methods for the Screening of Compounds
; TITLE OF INVENTION: Useful in the Treatment of Neurological Diseases
; FILE REFERENCE: CASE-03828
; CURRENT APPLICATION NUMBER: US/09/850,799
; PRIOR APPLICATION NUMBER: 2001-05-08
; PRIOR FILING DATE: EARLIER APPLICATION NUMBER: 09/360,779
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Patentm Ver. 2.0
```

```

; SEQ ID NO 1
; LENGTH: 1752
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (112)..(1131)
US-09-850-799-1

Alignment Scores:
Pred. No.: 2.9e-38 Length: 1752
Score: 481.00 Matches: 122
Percent Similarity: 44.73% Conservative: 18
Best Local Similarity: 38.98% Mismatches: 68
Query Match: 18.59% Indels: 105
DB: 10 Gaps: 10

US-09-902-772-4 (1-478) x US-09-850-799-1 (1-1752)
QY 224 ThrGlyAlaIleThrPheIlePheProAsnThrSerValTyrProGluAlaThrGlnArg 243
Db 333 ACGGCGGGGTATTCGGCACCCAGCGCAAGCTGACGTGCCCCCGCGCGTCCCCCCC 392
QY 244 IleThrThrArgProAspLeuProTyrGlnGlnAlaIleArgSerAlaTrpThrSerHis 263
Db 393 ATCTCCACCGCCCA-----ProGlnLeu-AspProTyrGlnIleLe 416
QY 264 SerHisProThrGlnSerLysAlaThrGlnProSerSerThrValProLysThrGlu 283
Db 417 AGCGATGACAGACAGCGGACCTCCAGCCCTGCTGATCAATGTAATCAACAGATCC 476
QY 284 AspGlnArg-----ProGlnLeu-AspProTyrGlnIleLe 295
Db 477 CGTCGAGATGCTGCTTTTAAGGAAGAGAGAGCCGACGCTGGGGCGCG-----CT 527
QY 295 UGlyProThrSerSerArgLeuAlaAsnProGlySerGlyGlnIleGlnLeuTrpGlnPh 315
Db 528 GAGCCCTCGCGTACAGAA-----GGCAGCGGACAGATCCAGTTGTGGCACTT 575
QY 315 eLeuLeuGlnLeuLeuSerAspSerSerAsnSerAsnCysIleThrTrpGlnGlyThrAs 335
Db 576 TCTACTGGAGCTGCTGGGAGACCGCGCGGATCGCTGCTGGTGGAGGCGGCCA 635
QY 335 nGlyGluPheLysMetTyrAspProAspGluValAlaIleArgTrpGlyGluArgLysSe 355
Db 636 CGGGAGATTCAGCTCAGACCGACCGACGAGAGTGGCGGAGCTGGGCGAGCGCAAG 695
QY 355 rLysProAsnMetAsnTyrAspLysLeuSerArgAlaLeuArgTyrTyrAspLysAs 375
Db 696 CAAAGCCCAATATGAATCACTGACAAAGCTAAATGCGAGCACTGCGTACTACTGACAA 755
QY 375 nIleMetThrLysValHisProProGluSerSerMetTyrLysTyrProSerAspLeuPr 395
Db 756 CATCATAGCAAGGTG----- 771
QY 395 oTyrMetSerSerTyrHisGlyLysArgTyrAlaTyrLysPheAspPheHisGlyIleAl 415
Db 772 -----CACGGCAAGCGGTACGCTTACCGCTTTGACTTTCCAGGGCGCTGCG 815
QY 415 aglAlaLeuGlnPro-----HisAlaHisPro----- 424
Db 816 ACAGGCTTGGCCAGCCACACCGCGGACGCCACGCGCGCTGCGCGCGCGAGCGGC 875
QY 424 ----- 424
Db 876 AGCCGCGCGCCAGATGGCGCACTTTACAGCTCCCGGCTGGTGGCTCCACTGCCCTT 935
QY 425 -----GlnLysMetAsnPheValAlaProHisProProAlaLeuProValThrSe 441
Db 936 CCGCGGCGCTCCAAATCAACCTTAATGAGAGCTGGCGGCGGTGGGCGCGCGCTGCTT 995
QY 441 rSerSerPhePheAlaAlaProAsnProTyrTyrPheSerPro----- 455
Db 455 -----
```

Db 996 CTCT-----TACTGGCTGGTCCACAGCCACGCCGCTGC 1031
QY 456 -----ThnglylylTyPrProAsnthrArgleuPro----- 466
Db 1032 CGCGCCAGCCGCTGCGCTTACCAACCCCGGGCTTGACAGCCGCCGCTTTGG 1091
QY 467 -AlaAlaHisMetProSerHisLeuGlyThrHyrrTyPr 478
Db 1092 CGCGGTGGCCGCGCTTGCATTTGGGGGTCAATTAT 1128
RESULT 7
US-09-920-300A-1716
; Sequence 1716, Application US/09920300A
; Patent No. US20020136728A1
; GENERAL INFORMATION:
; APPLICANT: King, Gordon E.
; APPLICANT: Meagher, Madeline Joy
; APPLICANT: Xu, Jiangchun
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.547
; CURRENT APPLICATION NUMBER: US/09/920,300A
; NUMBER OF SEQ ID NOS: 1789
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1716
; LENGTH: 2188
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-920-300A-1716
Alignment Scores:
Pred. No.: 1,82e-34 Length: 2188
Score: 444.00 Matches: 138
Percent Similarity: 38.36% Conservative: 45
Best Local Similarity: 28.93% Mismatches: 134
Query Match: 17.16% Indels: 161
Db: 10 Gaps: 16
US-09-902-772-4 (1-478) x US-09-920-300A-1716 (1-2188)
QY 122 GLAARGARValIleValProAlaAspProThrLeuTrpSerHisValArgIn 141
Db 491 CAGCGCGCGCTGGCATTCACAAAGAACCCCTGGCTGTGGAGTGAACAACAGGTATGCCAG 550
QY 142 TrpLeuGluTrpAlaValIleLysGluTrpGlyLeuProAspValAspIleLeuLeuPheGln 161
Db 551 TGGCTTCTGGGCGCACCAATGATTCAGTCTGTGAGCTGTAATCTGCAGAGCTTC--- 607
QY 162 AsnIleAspGlyLysGluLeuLysMetThrLysAspAspPheGlnArgLeuThrPro 181
Db 608 GGGATTAATGGCCAGATGCTGTGTACTCTGGCAAGAACGGCTTCTGGAGCTGGCAGCT 667
QY 182 SerTyPAsnAlaAspIleLeuLeuSerHisLeu----- 192
Db 668 GACTTGTGGGTGACATTCCTGGGAACATCTGGACAAATGATCAAGAAACCAAGAA 727
QY 193 -----HisTyLeuArg 196
Db 728 AAGACAGAAAGATCAATATGAGAAATTCACACCTCAGCTCCGTTCTCATTTGATTAA 787
QY 197 GluThr----- 198
Db 788 AGCAATACATTAGTTTGGACAGAGAGCGCCTATGAAATGACAGACAGAAATTAC 847
QY 199 -----ProLeu-ProHisLeuThrSerAspAsp 208
Db 848 CCCAAAGCGGCGCTCTGACAGACATGTGTCCGCGCTTCACACCCAGCGTACTACAGCTCT 907
QY 208 LAspLysAlaLeuGlnAsnSerProArgLeu----- 218
Db 908 GAGCAGAGATTTCAGATGTTCCTCCCAAGTCTCGGCTCAGCTCCGTCAGCGTCACTACTGC 967

QY 219 -MetHisAlaArgAsnThrGlyAlaThrPheIle-----Pn 231
Db 968 TCTGTAGTCAGACACTTCCACGAGCAGCACTCAATTGCTCACCAACATTTCTGGAGAT 1027
QY 231 eProAsnthrSerValIlyrProGluAlaThrGlnArgIleThrThrArgProAspLeuPr 251
Db 1028 CCCAAAGACAGACT---CCCTGAGAACGTGGCGACAGTTCTGAGACTCAGACTCC 1084
QY 251 oTyrgIuGln-----AlaArgSerAlaTrp----- 260
Db 1085 CTCTTCAGTCAGTGAACAGCCAGTGTCTCTGCTGATGTGACAGGGGTCTCTCTTC 1144
QY 261 -----ThrSerHisSerHisProThrGlnSe 269
Db 1145 GAGAGCTTCGAAAGTACTGACAGCAGTCTCTGCTCATTAAGCCAAACCAATGCTTTC 1204
QY 269 rLysAlaThrGlnProSerSerThrValProLysThrGluAspGlnArgProGln 289
Db 1205 AAGGATTACA-----TCCAAAGAGAGTAGTACCAGTGC-GACCAAGCCAAACCACTTAT 1257
QY 289 uAspProTyrgIleLeuGlyProThrSerSerArgLeuAlaAsnProGlySerGly 309
Db 1258 ACCTGAGCTGTGCTGGCGCGCTTACA-----GGAAGTGGACC 1296
QY 309 nIleGlnLeuTrpGlnPheLeuLeuGluLeuLeuSerAspSerSerAsnSerAsnCys11 329
Db 1297 TATTCACTGTGGCAGATTCTCTCTGAGCTGCTATCAGACAAATCTCGCAGCATTCAT 1356
QY 329 eThrTrpGluGlyThrAsnGlyLupHelysMetThrAspProAspGluValAlaArg 349
Db 1357 CAGCTGAGCTGAGCAGCAGATGGAGTTTAAAGTGGCCAGCCCGATGAGGTGGCCGCG 1416
QY 349 gTrpGlyGluArgLysSerLysProAsnMetAsnTyPAspLysLeuSerArgAlaLeuAr 369
Db 1417 GTGGGAAAGAGGAAAAATATAGCCCAAGATGAACTACGAAAGCTAGCGGGCGCTTAG 1476
QY 369 gTyTyTyTyPAspLysAsnIleMetThrLysValHisProProGluSerSerMetTy 389
Db 1477 CTACTATTACGACMAAGATCATCTCCACAAAGAG----- 1510
QY 389 sTyPProSerAspLeuProTyMetSerSerTyHisGlyLysArgTyPAlaTyPAsp 409
Db 1511 -----TCGGGAGCGCTAGCTGATCCGCTT 1536
QY 409 e-----AspPheHisGlyIleAla 416
Db 1537 CGTGTGCACTTCACAGACTGTCTGGGGTTACAGCCCGAGAACTGCAGCCATCTGTG-- 1594
QY 416 nAlaLeuGlnProHisAlaHisProGlnLysMetAsnPheValAlaProHisProAl 436
Db 1595 -GGCGTCCAGCCCGACAGGAGAGTGAAGTCCCGG-----GACCACCTTGAGCC 1644
QY 436 AluProValThrSerSerPhePheAlaIleProAsnProTyTrpAsnSerProth 456
Db 1645 GGCCCAAGGCTCTGAGTAGTGGAGGCCCATCTCTGACCACTCTCCGAGAGCCAG 1704
QY 456 rGlyGly-----IleTyPPr 461
Db 1705 GAAGGACAGATTGAATGTCCAGGAAGTGGCCAAAGAGCAGTGGCTTATTTGATTC 1764
QY 461 oAsn-----ThrArgLeuProAlaAlaHisMetProSerHis 473
Db 1765 CAAMCAGCGCTCTTTCAGCAGGCTGCTCTGTGGCAGCAACGGCAGC 1813
RESULT 8
US-10-033-528-1716
; Sequence 1716, Application US/10033528
; Patent No. US20020131971A1
; GENERAL INFORMATION:
; APPLICANT: King, Gordon E.
; APPLICANT: Meagher, Madeline Joy
; APPLICANT: Xu, Jiangchun

```

; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.547C1
; CURRENT APPLICATION NUMBER: US/10/033,528
; CURRENT FILING DATE: 2001-12-26
; NUMBER OF SEQ ID NOS: 1896
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1716
; LENGTH: 2188
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-033-528-1716

Alignment Scores:
Pred. No.: 1,82e-34 Length: 2188
Score: 444.00 Matches: 138
Percent Similarity: 38.36% Conservative: 45
Best Local Similarity: 28.93% Mismatches: 134
Query Match: 17.16% Indels: 161
Gaps: 16

US-09-902-772-4 (1-478) x US-10-033-528-1716 (1-2188)
QY 122 GLUARGARGVALLEVALPROALAASPRPROTHLEUTRPSERTHRASPHISVALARGGLN 141
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 491 CAGCGCGCCTGGCGCATGAGTCAAGTGTGTAACGTAATCTGCAGAGGTTG-- 607
QY 142 TTPLEUJUTPRALVALVALYSGLUTRGLYLEUPROASPRVALASPILEULEUPHEGLN 161
   :|||||:|||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 551 TGGCTTCTGGGGCCACCAATGAGTCAAGTGTGTAACGTAATCTGCAGAGGTTG-- 607
QY 162 ASNILEASPILYSGILEUCYSLYSMETHTRLYSASPSAPSPHEGLNARGLEUTHPRO 181
   :||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 608 GGCATGATATGGCCAGATGCTGTGTAACTTGGCAAGAACGCTTTCTGGAGCTGGCACT 667
QY 182 SETYRNAALASPILEULEUSERHSIELEU----- 192
   :||:|||||:|||||:|||||
Db 668 GACTTTGTGGTGACATCTCTGGGAACATCTGGACAAATGATCAAGAAACCAAGAA 727
QY 193 ----- 196
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 728 AAGCAGACAGATCAATGTGAAGAAATTCACACCTCACTCCGTTCTCATTTGATTAAC 787
QY 197 GLUTHR----- 198
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 788 AGCAATACATTAGGTTTGGACAGACAGACGGCGCCTATGGAATGCAGACAGAAATTAC 847
QY 199 ----- 208
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 848 CCCAAGAGCGGCTCCTGCAGACAGCATGTGTCGGCTCCACACCCAGCGTACCTCCTCT 907
QY 208 LASPLYALALEUGLNASBSERPROARGLEU----- 218
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 908 GAGCAGGAGTTTCAAGATGTTCCCAAGTCTGGGCTCACTCCGTCAGCGTACCTACTGC 967
QY 219 -METHSALALARGASNTHTGLYGLALATHRPHLEI-----PH 231
   :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 968 TCTGTGTCAGGAGCTCCAGAGCAGCACTTGAAATTTGGTCCACCAACATTCGGAGCT 1027
QY 231 EPROASNTHTSERVALYTRPROGLUALATHRGALNARGILETHTRHARGPROASLEUPR 251
   :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 1028 CCCAAGACCAACGACT---CCCTGAGAGAGGTTGCGACAGCTTCGAGAGCTCAGACTCC 1084
QY 251 CTYRGUGLN-----ALATARGSERALATRP----- 260
   :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 1085 CTCTCTCCAGTCTGGAACACCACTGCTCTGCTGGATGTGCAACGGGTTCTTCCTTC 1144
QY 261 ----- 269
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1145 GAGAGCTTGAAGATGACGTGACAGCACTCTCTGCTCATATAGCCAACCATCTCTTTC 1204
QY 269 TLYSALATHGLNPROSERSESRTHRVALPROLYSTHGLUASPIALNARGPROGLN 289

```

```

Db 1205 AAGGATTACA-----TCCAGAGAGAGTGACCAGTG--GAGCAAGGCAACAGATTAT 1257
QY 289 UASPRPTYGLINLEUGLIPROTHRSERSESRARGLEULASNPROHSISERGLYGL 309
   :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 1258 ACCTGACAGCTGTGCTGGCGGCTTCACA-----GGAAGTGGACC 1286
QY 309 NILEGLNLEUTPRINPHEULEULEULEUSERASPSERSEASNSERASANCYSII 329
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1297 TATTCAGCTGTGGCAGTTTCTCCGTGAGCTGCTATACAGCAAAATCTCCAGTATCAT 1356
QY 329 EHTTPGLUGLYTHRSNGLIUPHELYSMETHTHRASPROASPIULVALATARGAR 349
   :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 1357 CAGCTGAGCTGGAACGAGATGAGTTAACTGCCACCCCGATGAGTGGCCGCGG 1416
QY 349 GTTPGLIULARGLYSSERTLYSPROASMETASNTYTRASPLYLEUSERTARGALALEAR 369
   :|||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 1417 GTGGGAAAGAGAAATTAATTAATCCAAAGATACCTAGCAAGAGCTGACCGGGCTTACG 1476
QY 369 GTYTRTYRTPYASPIASNILEMETHTRLYSVALNHSIPROGLUSERSESRMETYRLY 389
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1477 CTACTATTAGCACAAGAACATCATCCACAGACG----- 1510
QY 389 STYRPROSERASPLEUPROTYRMETSESRTYRHSIGLYSARGYALATYRLYSRPH 409
   :|||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 1511 -----TCCGGGGAAGCGTACGTGACCGCTT 1536
QY 409 e-----ASPHENHSIGLYLEIAGI 416
   :|||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 1537 CGTGTGGACCTCCAGAACTTGTGGGGTTACGCCCGGAGAGATGCAAGCCATCTTG-- 1594
QY 416 NALALEUGLNPROHISALAHNHSIPROGLNLYSMETASNPHEVALALAPROHISPROAL 436
   :|||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 1595 -GGGCTGCAGCCGACACAGAGAGACTAGTGGCGG-----GACCACCTGAGACC 1644
QY 436 ALEUPROVALTHRSERSESRPHEALALAPROASNPOTYTRTPASNSERPROTH 456
   :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 1645 GGGCCGAGGCTCGTGACTGATGAGGAAGCCATCTGACCAAGCTGCTCCGAGGACCCAG 1704
QY 456 RGLYGLY-----ILETYRPR 461
   :|||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 1705 GAAAGCAGGATGAAATGTCCAGGAAAGTGGCCAAAGACAGTGGCTTATTCACATCC 1764
QY 461 OASN-----THARGLEUPROALALAHNHSERPROSERHS 473
   :|||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 1765 CAACCAAGCCTCTTGACACAGCTGCTCCTTGTGGACGACAGGCGAC 1813

RESULT 9
US-09-920-300A-1693
; Sequence 1693, Application US/09920300A
; Patent No. US20020136728A1
; GENERAL INFORMATION:
; APPLICANT: King, Gordon E.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Xu, Jianshun
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.547
; CURRENT APPLICATION NUMBER: US/09/920,300A
; CURRENT FILING DATE: 2001-07-31
; NUMBER OF SEQ ID NOS: 1789
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1693
; LENGTH: 2268
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-920-300A-1693

Alignment Scores:
Pred. No.: 1.92e-34 Length: 2268
Score: 444.00 Matches: 138
Percent Similarity: 38.36% Conservative: 45

```

Best Local Similarity: 28.93% Mismatches: 134
 Query Match: 17.16% Indels: 161
 DB: 10 Gaps: 16

US-09-902-772-4 (1-478) x US-09-920-300A-1693 (1-2268)

```

OY 122 GUAARGARValIleValProAlaAspProThrLeuTrpSerThrAspHisValArgGln 141
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
OY 571 CACGGCGCGCTGGGATTCAGAAAGAACCCCTGGCTGGTGAGTACAGCAACAGGTATGCCAG 630
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
OY 142 TRPLEUGLUTRPaIaValLysGluTrpGlyLeuProAspValAspIleuLeuPheGln 161
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
OY 631 TGGCTTCTGGGCCACCAATGATGATGCTGTGTGAGCTGATGCTGAGAGTTTC--- 687
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
OY 162 AsnIleAspGlyLysGluLeuGlyLysMetThrLysAspAspPheGlnArgLeuThrPro 181
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
OY 688 GGCATGAATGGCCAGATGCTGTGTACCTGGCAAGAACGCTTTCGAGCTGGCACCTP 747
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
OY 182 SerTrpAsnAlaAspIleuLeuSerHisLeu----- 192
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
OY 193 -----HisTrpLeuArg 196
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
OY 808 AAGACAGAAAGATCATATGAGAAATTCACACCTCACCTCCGTTCTCATTTGATTAC 867
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
OY 197 GluThr----- 198
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
OY 868 AGCAATATCATAGTTTGTGGACAGACAGCGCCCTATGAAATGCAGACACGAATTCAC 927
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
OY 199 -----ProLeu-ProHisLeuThrSerAspAsp 208
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
OY 928 CCCAAAGCGCGCTCTGGACAGACATGCTCGCGCTCCACACCCAGCTACTACCTCT 967
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
OY 208 LAspLysAlaLeuGlnAsnSerProArgLeu----- 218
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
OY 988 GAGCAGAGATTCAGATGTTCCCAAGTCTCGGCTCAGCTCCGTCAGCGTACCTACTG 1047
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
OY 219 -MethIleAlaArgAsnThrGlyLysAlaThrPheIle-----Ph 231
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
OY 1048 TCTGTAGTCAGTCTCCAGGACCACTGAATTTGCTCACCACCAATTCCTGGAGCT 1107
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
OY 231 eProAsnThrSerValTrpProGlnAlaThrGlnArgIleThrThrArgProAspLeuP 251
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
OY 1108 CCCAAAGACACAGACT---CCCTGAGAAACGGTGGGACAGCTTCAGAGCTCAGACTCC 1164
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
OY 251 oTrpGluGln-----AlaArgArgSerAlaTrp----- 260
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
OY 1165 CTCCTCAGTCTCGAAGACAGCGTCTGCTGCTGATGTGCACAGGGTCTCTCTTC 1224
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
OY 261 -----ThrSerHisSerHisProThrGlnSe 269
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
OY 1225 GAGAGCTTGGAAGAGTACGTACAGCGACTCTCTGCTCAATAAAGCAACCAATGCTTC 1284
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
OY 269 rTrpAlaThrGlnProSerSerSerThrValProLysThrGlnAspGlnArgProGln 289
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
OY 1285 AAGGATTTACA-----TCCAAGAGAGAGTGAACCCAGTG-GAGCAAGGCAACCACTTAT 1337
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
OY 289 uAspProTrpGlnIleLeuGlyProThrSerArgLeuAlaAsnProGlySerGly 309
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
OY 1338 ACCTGAGCTGTGCTGGCGGCTTCACA-----GGAAGTGGAC 1376
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
OY 309 nIleGlnLeuTrpGlnPheLeuLeuGlnLeuLeuSerAspSerSerAsnSerAsnGly 329
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
OY 1377 TATTTAGCTGTGGCACTTCTCTGAGCTGTATTCAGCAAAATCTTGACAGCATTTAT 1436
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
OY 329 eThrTrpGluGlyTrpAsnGlyLysPheLysMetThrAspProAspGlnValAlaArg 349
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
OY 1437 CAGCTGAGCTGGAAGAGAGTGGAGTTTAAGCTCGCCGACCCCGATGAGAGTGGCCGCG 1496
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
OY 349 gTrpGlyGluArgLysSerLysProAsnMetAsnTrpAspLysLeuSerArgAlaLeu 369
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
OY 1497 GTGGGAAAGAGAAATAATACCCCAAGATGAACATACGAAAGCTAGCGGGGCTTACG 1556
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

```

OY 369 gTrpTrpTrpAspLysAsnIleMetThrLysValHisProGluSerSerMetTrp 389
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
OY 1557 CTACTATTACAGACAGAAATCATTCACAAAGACG----- 1590
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
OY 389 sTrpProSerAspLeuProTrpMetSerSerTrpHisGlyLysArgTrpAlaTrp 409
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
OY 1591 -----TCGGGAAAGCGGTACTGATCGGCTT 1616
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
OY 409 e-----AspPheHisGlyIleAla 416
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
OY 1617 CGTGTGGACCTTCAGAACTTGCTGGGGTTACAGCCCGAGAACTGCAGCCATCTG--- 1674
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
OY 416 nAlaLeuGlnProHisAlaHisProGlnLysMetAsnPheValAlaProHisProAl 436
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
OY 1675 -GGCTTCACGCCCGACAGCAGGAGTGTAGTCCCGG-----GACCACCTCTBAGCC 1724
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
OY 436 AluProValThrSerSerSerPhePheAlaAlaProAsnProTrpTrpAsnSerPro 456
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
OY 1725 GGCCTCAGGCTGTGAGTGTAGTGGAAAGCCATCTTGACACGCTCTCCGAGACCCAG 1784
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
OY 456 rGlyGly-----IleTrpTrp 461
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
OY 1785 GAAAGCAGAGATTGAAATGTCCAGGAAAGTGCCCAAGAAAGAGTGCGCTTATTGCATCC 1844
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
OY 461 oAsn-----ThrArgLeuProAlaAlaHisMetProSerHis 473
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
OY 1845 CAAGCAGCGCTTGTGACAGCAGCTGCTCTTGTGGACAAAGCGCAC 1893
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

RESULT 10
 US-10-033-528-1693
 ; Sequence 1693: Application US/10033528
 ; Patent No. US20020131971A1
 ; GENERAL INFORMATION:
 ; APPLICANT: King, Gordon E.
 ; APPLICANT: Meagher, Madeleine Joy
 ; APPLICANT: Xu, Jiangchun
 ; APPLICANT: Secrist, Heather
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
 ; FILE REFERENCE: 210121.547C1
 ; CURRENT APPLICATION NUMBER: US/10/033, 528
 ; NUMBER OF SEQ ID NOS: 1896
 ; SOFTWARE: FASTSEQ for Windows Version 4.0
 ; SEQ ID NO 1693
 ; LENGTH: 2268
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; US-10-033-528-1693

Alignment Scores:
 Pred. No.: 1,92e-34 Length: 2268
 Score: 444.00 Matches: 138
 Percent Similarity: 38.36% Conservative: 45
 Best Local Similarity: 28.93% Mismatches: 134
 Query Match: 17.16% Indels: 161
 DB: 12 Gaps: 16

US-09-902-772-4 (1-478) x US-10-033-528-1693 (1-2268)

```

OY 122 GUAARGARValIleValProAlaAspProThrLeuTrpSerThrAspHisValArgGln 141
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
OY 571 CACGGCGCGCTGGGATTCAGAAAGAACCCCTGGCTGGTGAGTACAGCAACAGGTATGCCAG 630
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
OY 142 TRPLEUGLUTRPaIaValLysGluTrpGlyLeuProAspValAspIleuLeuPheGln 161
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
OY 631 TGGCTTCTGGGCCACCAATGATGATGCTGTGTGAGCTGATGCTGAGAGTTTC--- 687
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
OY 162 AsnIleAspGlyLysGluLeuGlyLysMetThrLysAspAspPheGlnArgLeuThrPro 181
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
OY 688 GGCATGAATGGCCAGATGCTGTGTACCTGGCAAGAACGCTTTCGAGCTGGCACCTP 747
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

```

QY 182 SerTyrAsnAlaSpIleLeuSerHisLeu-----192
Db 748 GACTTGTGGGTGACATTCCTGGGACATCTGGAGCAATGATCAAGAAACCAAGAA 807
QY 193 -----HisTyrLeuArg 196
Db 808 AAGCAGAAATGATATGAGAAATTCACACCTCAGCTCCCTCTCATTTGATTAAC 867
QY 197 GluThr-----198
Db 868 AGCAATACATTAGGTTTGGCAGACAGAGCCGCTATGGAATGCAGACAGAAATTAC 927
QY 199 -----ProLeu-ProHisLeuThrSerAspAspVa 208
Db 928 CCCAAGGCGGCTCTGGACAGACAGTGTCCGGCTCCACACCCAGCTCAGCTCT 987
QY 208 AspLysAlaLeuAlaSerProArgLeu-----218
Db 988 GAGCAGAGATTTCAGATGTTCCCAAGTCTCGGCTCAGCTCCGTCAGCTCAGCTAC 1047
QY 219 -MetHisAlaArgAsnThrGlyAlaThrPheIle-----Ph 231
Db 1048 TCTGTCACTGACGACTTCCAGCAGCAGCACTGAAATTGGTCACCAACAATTCGGACT 1107
QY 231 eProAsnThrSerValTyrProGluAlaThrGlnArgIleThrThrArgProAspLeuPr 251
Db 1108 CCCAAGACGACGACT---CCCTGAGACAGGTGCGACAGCTTCGAGAGCTCAGACTCC 1164
QY 251 oTyrGluGln-----AlaArgSerAlaTrp-----260
Db 1165 CTCTCCAGTCTGGAAACAGCCAGCTCCTGCTGATGTGCAAGCGGTTCTCTCTTC 1224
QY 261 -----ThrSerHisSerHisProThrGlnSe 269
Db 1225 GAGAGCTTCGAAGATGACTGACGACGACTCTCTGCTCATATAGCAACCATGCTTTC 1284
QY 269 rLysAlaThrGlnProSerSerThrValProLysThrGluAspGlnArgProGlnLe 289
Db 1285 AAGGATTAC-----TCCAGAGAGAGGAGTACCAGCTG- GAGCAGAGCAACACAGTTAT 1337
QY 289 uAspProTyrGlnIleLeuGlyProThrSerSerArgLeuAlaAsnProGlySerGly 309
Db 1338 ACCTGCGAGTGTGCTGCGCGCTTCACA-----GGAAGTGGACC 1376
QY 309 nIleGlnLeuTyrPheLeuLeuGluLeuSerAspSerSerAsnSerAsnCysI 329
Db 1377 TATTCACCTGTGGCAGTTTCTCCGAGCTGCTATCAGACAAATCCGCAAGTATCAT 1436
QY 329 eThrTrpGluGlyThrAsnGlyLupheLysMetThrAspProAspGluValAlaArg 349
Db 1437 CACCTGACCTGGAGACGATGGAGTTAAGCTGCCGCCGATGAGTGGCGCCCGC 1496
QY 349 gTTPGLIArgLysSerLysProAsnMetAsnTyrAspLysLeuSerArgAlaLeuArg 369
Db 1497 GTGGGGAAAGAGAAATTAAGCCCAAGATGACTACGAGAGAGTGAAGCGGGGCTTACG 1556
QY 369 gTyrTyrTyrAspLysAsnIleMetThrLysValHisProGluIleSerMetTyrL 389
Db 1557 CTACTATTACAGACAAGAAATCATCCACAAGAGC-----1590
QY 389 sTyrProSerAspLeuProTyrMetSerSerTyrHisGlyLysArgTyrAlaTyrLys 409
Db 1591 -----TCGGGGAAGCGCTACCTGTAACCCCTT 1616
QY 409 e-----AspPheHisGlyIleAlaG 416
Db 1617 CGTGTGGACCTCCAGAACTTGTGGGGTTCACGCCCGAGAGAACTGCACGCCATCCG-- 1674
QY 416 nAlaLeuGlnProHisAlaHisProGlnLysMetAsnPheValAlaProHisProAla 436
Db 1675 -GGGTCACAGCCGACAGCAGAGAGCTGAGGTGCGCG--GACACACCTGAGGCC 1724
QY 436 aLeuProValThrSerSerSerPhePheAlaAlaProAsnProTyrTrpAsnSerProth 456

```

```

Db 1725 GGGCCAGGCTGTGAGTGTAGTGGAAGCCCATCTGACAGCTCTCGAGGAGCCAG 1784
QY 456 rGlyIle-----IleTyrPr 461
Db 1785 GAAGGCGAGATGAGAAATGTCAGAGAAAGTGCCCAAGACAGTGGCTTATTTGATTC 1844
QY 461 oAsn-----ThrArgLeuProAlaAlaHisMetProSerHis 473
Db 1845 CAACACAGCGCTCTTGACACAGGCTGCTCTGTCGACAGCAGCGCAC 1893

RESULT 11
US-09-925-300-420
; Sequence 420, Application US/09925300
; Patent No. US20020151681A1
; GENERAL INFORMATION:
; APPLICANT: Craig Rosen,
; APPLICANT: Steve Ruben
; TITLE OR INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PAl01
; CURRENT APPLICATION NUMBER: US/09/925,300
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05988
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; NUMBER OF SEQ ID NOS: 1890
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 420
; LENGTH: 1884
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (56)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (283)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-300-420

Alignment Scores:
Pred. No.: 3,15e-33 Length: 1884
Score: 430.50 Matches: 138
Percent Similarity: 38.28% Conservative: 45
Best Local Similarity: 28.87% Mismatches: 134
Query Match: 16.63% Indels: 162
DB: 10 Gaps: 16

US-09-902-772-4 (1-478) x US-09-925-300-420 (1-1884)
QY 122 GluArgArgValIleAlaProAlaAspProThrLeuThrProSerThrAspHisValArgGln 141
Db 181 CAGCGGCGCTGGGCAATTCACAAAGAACCCCTGGCTGTGAGATGAGCAACAGATATGCCAG 240
QY 142 TrpLeuGlnTrpAlaValLysGluTyrGlyLeuProAspVal-AspIleLeuLeuPheG 161
Db 241 TGGCTTCTGCGGGCCACCAATGACTTCAGTCTGTGAACGTAATCTGCGAGAGTTC-- 298
QY 161 nAsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArgLeuThrPr 181
Db 299 -GGCATGATAGGCCAGATGTGCTGTAACTTGGCAAGAACGCTTCTGTGAGCTGGCACC 357
QY 181 oSerTyrAsnAlaSpIleLeuSerHisLeu-----192
Db 358 TGACTTTGTGGGTGACATTTCTCTGGGAACATCTGGACAAATGATCAAGAAACCAAGA 417
QY 193 -----HisTyrLeuArg 196
Db 418 AAGACAGAAATGATATGAGAAATTCACACCTCAGCTCCGTTCTCATTTGATTA 477
QY 196 gGluThr-----198

```

```

Db 478 CAGCATACATTAGTTTGGCAGACAGCGCCCTATGATGCACACAGAAATTA 537
Qy 199 -----Proleu-ProHisLeuThrSerAsp 208
Db 538 CCCCAAGGCGGCTCTGGACAGATGTCGCGCTCCACACACAGCTACTGACGTC 597
Qy 208 ALAspLysAlaLeuGlnAsnSerProArgLeu----- 218
Db 598 TGAGCGAGAGTTTCAGATGTTCCCAAGTCTCGGCTCAGCTCCGTCAGCGTCACCTAC 657
Qy 219 --MethAlaArgAsnThrGlyAlaThrPheIlePhe----- 231
Db 658 CTCGTGACGACGAGACTTCCAGCAGCACTGAAATTTCTCACCACAACTTCTGGGAC 717
Qy 232 --ProAsnThrSerValTyrProGluAlaThrGlnArgIleThrArgProAspLeu 251
Db 718 GCCCAAGACACGAGCT---CCCTGAGAACGGTGGGACAGCTTCGAGAGTCAGACTC 774
Qy 251 rotyrGluGln-----AlaArgSerAlaTrp----- 260
Db 775 CTTCTCCAGTCTCTGGAACGACGTCCTGCTGGATGCAACGGGTCCTTCCTT 834
Qy 261 -----ThrSerHisSerHisProThrGlns 269
Db 835 CGAGACCTTGAAGATGACTGCAGCCAGTCTCTGCTCCTCAATAGCCCAACATGCTTT 894
Qy 269 erLysAlaThrGlnProSerSerThrValProLysThrGluAspGlnArgProGlnL 289
Db 895 CAAGGATTCACA-----TCCAGAGAGAGTGCACCGGTG-GAGCAAGCAACCAAGTTA 947
Qy 289 euAspProtyrGlnIleLeuGlyProThrSerArgLeuAlaAsnProGlySerGly 309
Db 948 TACCTGACAGTGTGTCGGCGGCTTCACA-----GGAAGTGGAC 966
Qy 309 InIleGlnLeuTrpGlnPheLeuGlnLeuLeuSerAspSerSerAsnSerAspCysI 329
Db 987 CATATCAGCTGGGAGATTCTCTGAGAGTCTATCAGCAAAATCTGCACTATCA 1046
Qy 329 IeThrTrpGluGlyThrAsnGlyGluPheLysMetThrAspProAspGluAlaArg 349
Db 1047 TCAGCTGAGCTGAGACGCGATGGAGTTTAAAGTCCGCGCCGATGAGTGGCCGCG 1106
Qy 349 rGTpGlyGluArgLysSerLysProAsnMetAsnTyrAspLysLeuSerArgAlaLeu 369
Db 1107 GGTGGGGAAGAGAAATATAGCCCAAGTGAAGTACGAGAGCTGAGCGGGGCTTAC 1166
Qy 369 rGTyTyrTyrAspLysAsnIleMetThrLysValHisPropGluSerMetTyrL 389
Db 1167 GCTACTATTACGACAAACATCATCCACAGAGC----- 1201
Qy 389 ySTyTyrProSerAspLeuProTyrMetSerSerTyrHisGlyLysArgTyrAlaTyrLysP 409
Db 1202 -----TCGGGAGAGCCCTTCAGTTCACCGCT 1226
Qy 409 he-----AspPheHisGlyIleAlaG 416
Db 1227 TCGTGTGCACTCCAGAACTGCTGGGGTTCAGGCCCGAGAACGTGACGCGCATCTCG- 1285
Qy 416 InAlaLeuGlnProHisAlaHisProGlnLysMetAsnPheValAlaProAsnProPro 436
Db 1286 --GGGCTCCAGCCCGACAGGAGAGTGAAGTGGCGG-----GACCAACCTTGAGC 1334
Qy 436 lAlaLeuProValThrSerSerSerPhePheAlaAlaProAsnProTyrTyrPasnSerProT 456
Db 1335 CGGCCCCAGGCTGCTGGAGTGAAGCCCATCTGACCGACGTCGCCGAGACCA 1394
Qy 456 hGlyGly-----IleTyrP 461
Db 1395 GGAAGGCGAGATTGAATGTCCAGAAAGTGGCCAGAGACAGTGGCCCTTAATGTCATC 1454
Qy 461 roAsn-----ThrArgLeuProAlaAlaHisMetProSerHis 473
Db 1455 CCAAAACCAAGCGCTTGTACAGAGCTGCTCCCTGTGGCAGCAGACGGCAC 1504

```

```

RESULT 12
US-09-864-761-29364/C
; Sequence 29364, Application us/09864761
; Patent No. US20020048763A1
GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Aecmicha X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 29364
; LENGTH: 348
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AP001426.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.73
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.52
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.45
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.55
; OTHER INFORMATION: SWISSPROT HIT: P11308, EVALUATE 4.00e-38
; OTHER INFORMATION: NT HIT: M21535.1, EVALUATE 1.00e-113
; OTHER INFORMATION: EST_HUMAN HIT: AW948986.1, EVALUATE 1.00e-108
US-09-864-761-29364
Alignment Scores:
Pred. No.: 1,61e-29 Length: 348
Score: 382.00 Matches: 69
Percent Similarity: 98.63% Conservative: 3
Best Local Similarity: 94.52% Mismatches: 1

```

```
Query Match: 14.76% Indels: 0
DB: 10 Gaps: 0
US-09-902-772-4 (1-478) x US-09-864-761-29364 (1-348)
QY 129 AlaSPProThrLeuTrpSerThrAspHisValArgGlnTrpLeuGluTrpAlaValIys 148
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
DB 313 TCAGATCCTACGCTATGAGTACAGACCATGTGGCGAGTGGCTGAGACTGGCGGTGAA 254
QY 149 GluTyrGlyLeuProAspValAspIleLeuLeuPheGlnAsnIleAspGlyLysGluLeu 168
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
DB 253 GAATATGGCCCTTCCAGAGCTCAACATCTGTATTCCGAACATCGATGGAGAACTG 194
QY 169 CysLysMetThrLysAspAspPheGlnArgLeuThrProSerTyrAsnAlaAspIleLeu 188
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
DB 193 TGCAAGATGACCAAGAGAGACTTCCAGAGGCTCAACCCCACTCAACAGCCGACATCCTT 134
QY 189 LeuSerHisLeuHisTyrLeuArgGluThrProLeuPro 201
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
DB 133 CTCACATCTCCACTACTCTCAGAGAGTAGTAACTCCCC 95

RESULT 13
US-09-864-761-5035/c
: Sequence 5035, Application US/09864761
: Patent No. US20020048763A1
: GENERAL INFORMATION:
: APPLICANT: Penn, Sharon G.
: APPLICANT: Rank, David R.
: APPLICANT: Hanzel, David K.
: APPLICANT: Chen, Wensheng
: TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
: FILE REFERENCE: Aeomica-X-1
: CURRENT APPLICATION NUMBER: US/09/864,761
: CURRENT FILING DATE: 2001-05-23
: PRIOR APPLICATION NUMBER: US 60/180,312
: PRIOR FILING DATE: 2000-02-04
: PRIOR APPLICATION NUMBER: US 60/207,456
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: US 09/632,366
: PRIOR FILING DATE: 2000-08-03
: PRIOR APPLICATION NUMBER: GB 24263,6
: PRIOR FILING DATE: 2000-10-04
: PRIOR APPLICATION NUMBER: US 60/236,359
: PRIOR FILING DATE: 2000-09-27
: PRIOR APPLICATION NUMBER: PCT/US01/00666
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00667
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00664
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00669
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00665
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00668
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00663
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00662
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00661
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00670
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: US 60/234,687
: PRIOR FILING DATE: 2000-09-21
: PRIOR APPLICATION NUMBER: US 09/608,408
: PRIOR FILING DATE: 2000-06-30
: PRIOR APPLICATION NUMBER: US 09/774,203
: PRIOR FILING DATE: 2001-01-29
: NUMBER OF SEQ ID NOS: 49117
: SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
```

```
: SEQ ID NO 5035
: LENGTH: 477
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: OTHER INFORMATION: MAP TO AP000021.2
: OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.95
US-09-864-761-5035

Alignment Scores:
Pred. No.: 2.56e-29 Length: 477
Score: 382.00 Matches: 69
Percent Similarity: 98.63% Conservative: 3
Best Local Similarity: 94.52% Mismatches: 1
Query Match: 14.76% Indels: 0
DB: 10 Gaps: 0

US-09-902-772-4 (1-478) x US-09-864-761-5035 (1-477)
QY 129 AlaSPProThrLeuTrpSerThrAspHisValArgGlnTrpLeuGluTrpAlaValIys 148
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
DB 454 TCAGATCCTACGCTATGAGTACAGACCATGTGGCGAGTGGCTGAGTGGCGGTGAA 395
QY 149 GluTyrGlyLeuProAspValAspIleLeuLeuPheGlnAsnIleAspGlyLysGluLeu 168
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
DB 394 GAATATGGCCCTTCCAGAGCTCAACATCTGTATTCCGAACATCGATGGAGAACTG 335
QY 169 CysLysMetThrLysAspAspPheGlnArgLeuThrProSerTyrAsnAlaAspIleLeu 188
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
DB 334 TGCAAGATGACCAAGAGAGACTTCCAGAGGCTCAACCCCACTCAACAGCCGACATCCTT 275
QY 189 LeuSerHisLeuHisTyrLeuArgGluThrProLeuPro 201
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
DB 274 CTCACATCTCCACTACTCTCAGAGAGTAGTAACTCCCC 236

RESULT 14
US-09-864-761-12798/c
: Sequence 12798, Application US/09864761
: Patent No. US20020048763A1
: GENERAL INFORMATION:
: APPLICANT: Penn, Sharon G.
: APPLICANT: Rank, David R.
: APPLICANT: Hanzel, David K.
: APPLICANT: Chen, Wensheng
: TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
: FILE REFERENCE: Aeomica-X-1
: CURRENT APPLICATION NUMBER: US/09/864,761
: CURRENT FILING DATE: 2001-05-23
: PRIOR APPLICATION NUMBER: US 60/180,312
: PRIOR FILING DATE: 2000-02-04
: PRIOR APPLICATION NUMBER: US 60/207,456
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: US 09/632,366
: PRIOR FILING DATE: 2000-08-03
: PRIOR APPLICATION NUMBER: GB 24263,6
: PRIOR FILING DATE: 2000-10-04
: PRIOR APPLICATION NUMBER: US 60/236,359
: PRIOR FILING DATE: 2000-09-27
: PRIOR APPLICATION NUMBER: PCT/US01/00666
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00667
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00664
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00669
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00665
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00668
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00663
: PRIOR FILING DATE: 2001-01-30
```

```

; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 12798
; LENGTH: 519
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AP001426.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.73
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.52
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.45
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.55
; US-09-864-761-12798

```

```

Alignment Scores:
Pred. No.: 2,9e-29 Length: 519
Score: 382.00 Matches: 69
Percent Similarity: 98.63% Conservative: 3
Best Local Similarity: 94.52% Mismatches: 1
Query Match: 14.76% Indels: 0
DB: 10 Gaps: 0

```

US-09-902-772-4 (1-478) x US-09-864-761-12798 (1-519)

```

QY 129 AlaSPProThleuTrpserThrasphsValArgGlnTrpleuGluTrpAlaValLys 148
; ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
DB 360 TCAGATCCACGCTATGAGTACAGACCATGTCCGCGAGTGGAGTGGCGGTGAAA 301
QY 149 GluTyGlyLeuProAspAlaPheLeuLeuPheGlnAsnLleAspGlyLysGluLeu 168
; ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
DB 300 GAATATGGCTTCACAGCGTCACATCTGTATTCTCAGAACTCATGGAGAACTG 241
QY 169 CysLysMetThrLysAspAspPheGlnArgLeuThrProSerTyzAsnAlaSpLleLeu 188
; ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
DB 240 TCGAAGATGACCAAGACGACTCCAGAGGCTCACCCCGACGTACAAAGCCGACATCCTT 181
QY 189 LeuSerHisLeuHisTyrlleuArgGluThrProLeuPro 201
; ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
DB 180 CTCCTCATCTCCACTCACTCCTCAGAGAGAGTAGCTCC 142

```

RESULT 15

```

; US-09-864-761-7427/c
; Sequence 7427, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecm1ca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6

```

```

; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 7427
; LENGTH: 553
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AP001426.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.76
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2
; US-09-864-761-7427

```

```

Alignment Scores:
Pred. No.: 1,96e-28 Length: 553
Score: 374.00 Matches: 71
Percent Similarity: 100.00% Conservative: 4
Best Local Similarity: 94.67% Mismatches: 0
Query Match: 14.45% Indels: 0
DB: 10 Gaps: 0

```

US-09-902-772-4 (1-478) x US-09-864-761-7427 (1-553)

```

QY 5 IleLysGluAlaLeuSerValValSerGluAspGlnSerLeuPheGluCysAlaTyrgly 24
; ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
DB 297 ATTCAGAGAGCCCTATCACTTGTGAGTGAAGACCATCTGTTTGAAGTGGCTACGGA 238
QY 25 SerProHisLeuAlaLysThrGluMetThrAlaSerSerSerGluTyrglyGlnThr 44
; ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
DB 237 ACGCCACACCTGTGATAGACAGATGATGACCGCGTCTCTCCACGACATATGACAGACT 178
QY 45 SerLysMetSerProArgValProGlnGlnAspTrpLeuSerGlnProProAlaArgVal 64
; ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
DB 177 TCCAGATGAGCCACGCGCTCCCTCCAGAGATGGCTGTCTCAACCCCGACGAGGTC 118
QY 65 ThrLysMetGluCysAsnProAsnGlnValAsnGlySerArg 79
; ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
DB 117 ACCATCAAAATGAATGTAACCTTACGACGAGTGAATGCTCAAG 73

```

Tue Nov 12 09:10:45 2002

us-09-902-772-4.rnpb

Page 13

Search completed: November 9, 2002, 21:28:47
Job time : 78.9473 secs

THIS PAGE BLANK (USPTO)

Gencore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 9, 2002, 17:03:37 ; Search time 1595.56 Seconds

(without alignments)
4851.857 Million cell updates/sec

Title: US-09-902-772-4

Perfect score: 2588

Sequence: 1 MASTIKREALSVESDQSLFE.....IYPNRLPAAMPSHLGTTY 478

Scoring table:

BLOSUM62	Xgapop 10.0 , Xgapext 0.5
Xgapop 10.0 , Xgapext 0.5	
Xgapop 6.0 , Xgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xip
-O/cgnt2_1/USPTO.spool/US09902772/runat.07112002.134034.9041/app.query.fasta.1.1294
-DB=EST -OEMT=fastap -SUFFIX=prt -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=BIOSUM62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -NORM=ext -HEAPSIZE=500 -MINLEN=0 -ALIGN=15 -MODE=LOCAL
-USPR=US09902772.gcgn.1.1.2221.runat.07112002.134034.9041 -NCPU=6 -ICPU=3
-NO_XLPHY -NO_MMAPP -LARGEDUPLY -NEG_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-WANT.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estlun:*
5: em_estlov:*
6: em_estlpi:*
7: em_estro:*
8: em_htc:*
9: gb_estli:*
10: gb_estli2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estlom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_lin:*
20: em_gss_pln:*
21: em_gss_vit:*
22: em_gss_fun:*
23: em_gss_man:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1153	44.6	1045	13	BM456833
2	1141.5	44.1	880	14	BQ233264
3	1121	43.3	634	14	BM489636
4	1111.5	42.9	1089	14	BQ212127
5	1085	41.9	915	12	BG390291
6	1053	40.7	715	12	BG388025
7	1010	39.0	629	10	AM948986
8	992.5	38.4	928	12	BG259459
9	988	38.2	701	9	AU136709
10	984	38.0	642	13	BI713036
11	984	38.0	668	10	BB660034
12	980	37.9	592	13	BM088063
13	968	37.4	827	9	AJ456498
14	948	36.6	675	10	BB639043
15	932	36.0	629	10	BB577887
16	882.5	34.1	1074	14	BQ953977
17	868.5	33.6	520	10	BE233460
18	845.5	32.7	949	13	BI558012
19	806	31.1	1015	9	AL554108
20	793	30.6	522	12	BF566705
21	769	29.7	516	12	BG691566
22	721	27.9	645	10	BB656030
23	715	27.6	666	12	BG390572
24	707	27.3	619	10	AM421310
25	701	27.1	671	13	BI558530
26	691	26.7	553	12	BG256864
27	686.5	26.5	524	9	AJ794504
28	666	25.7	1245	14	BQ212133
29	653.5	25.3	1068	14	BQ062565
30	648.5	25.1	872	14	BQ232274
31	643.5	24.9	740	9	AJ456506
32	639	24.7	894	14	BQ439561
33	619	23.9	815	9	AJ446143
34	616.5	23.8	756	9	AJ445352
35	609	23.5	787	10	BE559913
36	600.5	23.2	874	14	BQ215477
37	597.5	23.1	947	14	BQ642921
38	596.5	23.0	587	9	AL598619
39	588	22.7	651	10	BB442803
40	581.5	22.5	770	12	BG701574
41	580	22.4	788	9	AJ456869
42	573	22.1	907	10	BE561128
43	567.5	21.9	810	10	BE560516
44	558.5	21.6	972	12	BG257780
45	556	21.5	674	10	BB238163

ALIGNMENTS

RESULT 1
BM456833
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE

BM456833 1045 bp mRNA linear EST 05-FEB-2002
AGENCOURT 6404041 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:5583753
5', mRNA sequence.
BM456833
BM456833.1 GI:18505873
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 1045)
NIH-MGC <http://mgc.ncl.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Life Technologies, Inc.
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM12347 row: f column: 10
 High quality sequence stop: 669.

FEATURES
 source Location/Qualifiers

1..1045
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5583753"
 /clone_lib="NIH_MGC_92"
 /tissue_type="embryonal carcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: testis; Vector: pCMV-SPORT6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
 Average insert size 2.5 kb. Library enriched for
 full-length clones and constructed by Life Technologies.
 Note: this is a NIH_MGC Library."

BASE COUNT 300 a 290 c 230 g 225 t
 ORIGIN

Alignment Scores:

Pred. No.: 3,54e-106 Length: 1045
 Score: 1153.00 Matches: 236
 Percent Similarity: 76.078 Conservative: 12
 Best Local Similarity: 72.398 Mismatches: 24
 Query Match: 44.554 Indels: 55
 DB: 13 Gaps: 7

US-09-902-772-4 (1-478) x BM456833 (1-1045)

QY 7 GUAUAlaUeUSeRValValSeRgIuaSpGInSeRleuPheGInUcYsaLaTyRgIySeRPro 26
 DB 190 GAAGCCTTATCACTGTGAGTGAAGACCAAGTCGTTGAGTGCGCTACGGAACGCCA 249
 QY 27 HIsleUaIaUyShrgIuMeThraIaSeRSeRSeRSeRSeRSeRSeRSeRSeRSeRSeRSeR 46
 DB 250 CACCTGGCTAAAGACAGAGAGACCGGCTCTCTCCAGCAGCATATGACGACACTTCCAG 309
 QY 47 MeSeRProArYalYProGInGInaSPTrPLeuSeRgInProBProAlaArGValThrIle 66
 DB 310 ATGAGCGCAGCGCTCCCTCAGCAGATGGCTGTCTCAACCCCGCAGCGACGACCATC 369
 QY 67 LysMeTGlucYsaNProaSnGInValaSnGlySeRArgaNsSeRProaSpaSPCySseR 86
 DB 370 AAAATGAATGTAACCTTCAAGCCAGGTGAATGGCTCAAGGAACCTCTCATGATGAGT 429
 QY 87 ValAlaIySGlyGlyLysMeTValSeRSeRSeRSeRSeRSeRSeRSeRSeRSeRSeRSeR 106
 DB 430 GTGGCCAAAGGCGGGAAGATGGTGGCAGCCAGACACCGTTGGAGTGAACCTGAGGAGC 489
 QY 107 TyMeTGlUGlUyShISlePProPProaSPMeTThrThraSnGInUaRGAATValIle 126
 DB 490 TACATGAGAGAGACACATGACCCCGCAACATGACCCAGACGACGACGACGACGACG 549
 QY 127 ValProAlaSPProThrIeUTrPSeRThraSPHisValaRgInTrPLeuGInTrPAla 146
 DB 550 GTGCCAGCAAGATCTTACGATATGAGATACAGACCATGTGGGAGTGGAGTGGGCG 609
 QY 147 ValUyGInUTrGlyLeuProaSPValaSPlleUleuLeuPheGInaSnIleasPglyLys 166
 DB 610 GTGAAGAATATGGCTTCAGACGTCACATCTTGTATTCAGAAACATCGATGGGAG 669
 QY 167 GlUeUcYsIySeMThrIySaSPaSPPhelArgLeuThrProSeRTyRaaAlaSP 186
 DB 670 GAACGTGTGCAAGATGACCAAGACGACTTCCAGAGGCTCACCCCGACGTACAAAGCCGAC 729

QY 187 ILeuLeuSeRHisIeUHisTyRLeuArgGInThrProLeuProHisIeUThrSeRAsp 206
 DB 730 ATCTTCTCTCACATCTTCACTACCTCAGAGACTCTCTTCCACTTGTGACTCCAGAT 789
 QY 207 AsPValaSPySaIaUeUInaSnSeRProArIeUSeMThraIaArGaNsTrNGlyGly 226
 DB 790 GATGTGATTAAGCTTACAACTCTCCCGGTTAAATGATCTGAAACACC----- 842
 QY 227 AlaThrPheIlePheProaSnThrSeRValTyPProGInaIaThrgInArgIleThrTh 246
 DB 842 ----- 842
 QY 247 ArgProaSPleuProTyRgInGInAlaArgArSerAlaTrPThrSeRHisSeRPro 266
 DB 843 -----GATTTACCATAATGAGACCCCGCGGATGACCGCTGACAGGTCAACCC 896
 QY 267 ThrGInSeRySaIaThrgInProSeRSeRThraI-----ProLys 281
 DB 897 -----CAAGCCCAAGTCGAAGAGCTGTTCAAACATTTCTCCCTTCACAAAGGCCAAA 947
 QY 282 ThrGInaSPGInaRgProGInLeuAsPProTyRgInIleUeUgIyProThrSeRArg 301
 DB 948 ACTGAA-----AGACCCAGC-----GCCCTCAGGT----- 974
 QY 302 LeuAlaSaNProGlySeRgInIleGInLeuTrPcInPheUeUgInUleuLeuSeR 321
 DB 975 ---TAAATATCC-----TTAAATCAACAATCTTGG 1001
 QY 322 AsPSeRSeRSeRSeRSeR 327
 DB 1002 GAACCCCAATATTAAC 1019

RESULT 2
 LOCUS BQ233264 880 bp mRNA linear EST 02-MAY-2002
 DEFINITION AGENCOURT_7565836 NIH_MGC_92 Homo sapiens cDNA IMAGE:6041526
 5', mRNA sequence.
 ACCESSION BQ233264
 VERSION BQ233264.1 GI:20414664
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 880)
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Life Technologies, Inc.
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM13279 row: h column: 07
 High quality sequence stop: 608.

FEATURES source

1..880
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:6041526"
 /clone_lib="NIH_MGC_92"
 /tissue_type="embryonal carcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: testis; Vector: pCMV-SPORT6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
 Average insert size 2.5 kb. Library enriched for
 full-length clones and constructed by Life Technologies.
 Note: this is a NIH_MGC Library."

```

BASE COUNT      250 a      231 c      207 g      189 t      3 others
ORIGIN

Alignment Scores:
Pred. No.:      4e-105      Length:      880
Score:          1141.50     Matches:     221
Percent Similarity: 95.02%   Conservative: 8
Best Local Similarity: 91.70% Mismatches:    8
Query Match:    44.11%      Indels:      4
DB:             14          Gaps:        1

US-09-902-772-4 (1-478) x BQ233264 (1-880)

QY      1 MetAlaSerThrIleLysGluAlaLeuSerValValSerGluAspGlnSerLeuPheGlu 20
      |||
Db      74 ATGCCCGACGACTATTAAGAAAGCCCTTATCAGTGTGAGTACAGACACGATCGTTGTTCAG 133
QY      21 CysAlaIleArgSerProHisLeuAlaLysThrGluMetThrAlaSerSerSerGlu 40
      |||
Db      134 TGTGCTTACGGAGACGCCACCTGGCTTAAGACAGATGACCGGCTCTCTCCAGCGAC 193
QY      41 TyrGlnIleThrSerLysMetSerProArgValProGlnGlnAspTyrLeuSerGlnPro 60
      |||
Db      194 TATGGACAGACTTCCAAAGATGAGCCACGGCTCCCTCAGCAGATTTGGCTGTCTCAACCC 253
QY      61 ProAlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsn 80
      |||
Db      254 CCACCCAGGGGACACATCAAAATGATGTAACCTTACGAGCGATGTAATGGCTCAAGGAC 313
QY      81 SerProAspAspCysSerValAlaLysGlyLysMetValSerSerSerAspAsnVal 100
      |||
Db      314 TCTCTCTATGATATCAATGTGGCCAAAGGGGGAAGTGGGGGACCCACGACACCGTT 373
QY      101 GlyMetAsnTyrGlySerTyrMetGluGluLysHisIleProProProAsnMetThrThr 120
      |||
Db      374 GGGATGAACTACGGCAGCTTACATGAGAGAGACATGCCACCCCAACATACACACG 433
QY      121 AsnGluArgArgValIleValProAlaAspProThrLeuTyrSerThrAspHisValArg 140
      |||
Db      434 AACGAGCGCAGACTTATCGTCCAGACATCCATGAGTACAGACATGTCGG 493
QY      141 GluTyrPheGluTyrPalaValLysGluTyrGlyLeuProAspValAspIleLeuPhe 160
      |||
Db      494 CAGTGGCTGAGTGGGGGGAAGATATGGCTTCCAGACGTCACATCTTGTATTC 553
QY      161 GlnAsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArgLeuThr 180
      |||
Db      554 CAGAAATCATGAGGAAGAACTGTSCAAGATGACCAAGAGCATCTCCAGAGGCTCAC 613
QY      181 ProSerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGluThrProLeu 200
      |||
Db      614 CCCAGTACAAACCGCAGACATCTCTCTCACAATCTCCATACCTCAGAGAGACTCGTCT 673
QY      201 ProHisLeuThrSerAspAspValAspLysAla-LeuGlnAsnSerProArg-LeuMet 220
      |||
Db      674 CCACATTTGACTTCAGATGATGATGATGAAGCCCTTACAAACTCTCCAGGGGTTAATGC 733
QY      220 1s-AlaArgAsnThrGlyAlaThrPheIlePheProAsn---ThrSerValTyr 237
      |||
Db      734 ATGCTAGAAACACAGAGGGGTGACGCTTTTATTTTCCAAATTAACCTCAGTAAT 790

RESULT 3
LOCUS      BM489636      634 bp      mRNA      linear      EST 07-FEB-2002
DEFINITION pgm2n.pk011.124 Normalized Chicken Breast Muscle, Leg Muscle, and
            Epiphyseal Growth Plate cdna library (pgm2n) Gallus gallus cdna
            clone pgm2n.pk011.124 5' similar to sp|Q90837|ERG.CHICK
            TRANSCRIPTIOMAL REGULATOR ERG p1r1560754 transcription factor erg -
            chicken emb|CA54404.1| (X71159) ERG [Gallus gallus], mRNA
            sequence.
ACCESSION  BM489636
VERSION    BM489636.1  GI:18610567
KEYWORDS   EST.

```

```

SOURCE      chicken.
ORGANISM    Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1 (bases 1 to 634).
AUTHORS    Cogburn, J.A. and Monsonego-Ornan, E.
TITLES      ESTs from Normalized Chicken Breast Muscle, Leg Muscle, and
            Epiphyseal Growth Plate cdna library, USDA/IRAFs Animal Genome
            Project
JOURNAL     Unpublished (2002)
COMMENT     Contact: Larry A. Cogburn
            University of Delaware
            Townsend Hall, Newark, DE 19717, USA
            Tel: 302-831-1335
            Fax: 302-831-2822
            Email: cogburn@udel.edu, www.chickest.udel.edu.

FEATURES
            source
            Location/Qualifiers
            1..634
               /organism="Gallus gallus"
               /strain="Commercial broiler and Ottawa Res. Centre
               strains 90 & 21"
               /db_xref="taxon:9031"
               /clone="pgm2n.pk011.124"
               /clone_lib="Normalized Chicken Breast Muscle, Leg Muscle,
               and Epiphyseal Growth Plate cdna library (pgm2n)"
               /sex="Male and Female"
               /tissue_type="Breast muscle, leg muscle and epiphyseal
               growth plate"
               /dev_stage="Breast, leg; Embryo(d19); post-hatch(1d, 1, 3, 5, 7, 9
               , 11 weeks); growth plate(1d, 7d, 14d post-hatch)"
               /lab_host="E. coli EMD10B"
               /note="Vector: PCWVS-PORT6; library made from equivalent
               pools of total RNA isolated from each tissue (embryonic
               muscle 33.3%, juvenile muscle 33.3%, and epiphyseal growth
               plate 33.3% of the final RNA pool). Single pass sequencing
               from 5'-end"
BASE COUNT  192 a      164 c      138 g      140 t
ORIGIN

Alignment Scores:
Pred. No.:      2.92e-103      Length:      634
Score:          1121.00     Matches:     210
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00% Mismatches:    0
Query Match:    43.32%      Indels:      0
DB:             13          Gaps:        0

US-09-902-772-4 (1-478) x BM489636 (1-634)

QY      120 ThrAsnGluArgArgValIleValProAlaAspProThrLeuTyrSerThrAspHisVal 139
      |||
Db      3 ACCAATGACGAGAGATTAATTTGTCACAGATCTTACGTTATGACACAGACAGACATGTA 62
QY      140 ArgGlnTyrPheGluTyrPalaValLysGluTyrGlyLeuProAspValAspIleLeuLeu 159
      |||
Db      63 CGGCACTGGCTGGAGTGGGAGTGAAGAGATGCTTCCACAGACGAGGACATCTTTTG 122
QY      160 PheGlnAsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArgLeu 179
      |||
Db      123 TTCAGAAACATGTGATGGGAAAGAGTTGTGTAAATGACCAAAAGATGACTTCCAGAGACTC 182
QY      180 ThrProSerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGluThrPro 199
      |||
Db      183 ACCGCGAGCTATTAACGAGATATCTCTCTGACACCTTACACTTACAGAGAGACTCT 242
QY      200 LeuProHisLeuThrSerAspAspValAspLysAlaLeuGlnAsnSerProArgLeuMet 219
      |||
Db      243 CTTCACATTTGACTTCAGATGATGATGATGAAGCCCTTACAAACTCTCCACGGTTAATG 302
QY      220 HisAlaArgAsnThrGlyAlaThrPheIlePheProAsnThrSerValTyrProGlu 239
      |||
Db      303 CATGCTAGAAACACAGAGAGACCATTTTATTTTCCAAATATCATCAGTTTACCCAGAA 362

```

QY	240	AlaThrGlnArgIleThrThrArgProAspLeuProTylGluGlnAlaArgSerAla	259		
Db	363	GGAAGCAAGATATAACAACAAGGCCAGATTACCTTATGACCAAGCGAGATACGG	422		
QY	260	TrpThrSerHisSerHisProThrGlnSerLysAlaThrGlnProSerSerSerThrVal	279		
Db	423	TGGAGAGTCAACAGCCATCCACTGACAGTCAAAAGCTACCCAAACATCATCTTCAACAGTg	482		
QY	280	ProLysThrGlnAspGlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSer	299		
Db	483	CCCAAAACGAAGACCAAGCGTCTCTGATTGATACCTTATCAATTCCTTGGACCGACACg	542		
QY	300	SerArgLeuAlaAsnProGlySerGlyGlnIleGlnLeuTrpGlnPheLeuLeuGlnLeu	319		
Db	543	AACCGCTTGCAGAAATCCAGAGGAGTGGGCAAGATACAGCTATGGCAGTCTACTGGAGCTT	602		
QY	320	LeuSerAspSerSerAsnSerAsnGlySer	329		
Db	603	CTGTGGACAGCTCCAACTCCAACTGCATC	632		
RESULT 4	BQ212127	1089 bp	mRNA	linear	EST 02-MAY-2007
LOCUS	BQ212127				
DEFINITION	AGNCOCOURT_7571287 NIH_MGC_92 Homo sapiens	cdna	clone	IMAGE:6052136	
ACCESSION	BQ212127				
VERSION	BQ212127.1	GI:20392058			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.				
TITLE	1 (bases 1 to 1089)				
JOURNAL	NIH-MGC http://mgc.nci.nih.gov/ .				
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC)				
	Unpublished (1999)				
	Contact: Robert Strausberg, Ph.D.				
	Email: cgapbs-r@mail.nih.gov				
	Tissue Procurement: ATCC				
	CDNA Library Preparation: Life Technologies, Inc.				
	CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)				
	DNA Sequencing by: Agencourt Bioscience Corporation				
	Clone distribution: MGC clone distribution information can be				
	found through the I.M.A.G.E. Consortium/LNL at:				
	http://image.lnl.gov				
	Plate: L1AM13307 row: b column: 09				
	High quality sequence stop: 562.				
FEATURES	Location/Qualifiers				
source	1..1089				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	/clone="IMAGE:6052136"				
	/clone_lib="NIH_MGC_92"				
	/tissue_type="embryonal carcinoma, cell line"				
	/lab_host="DH10B (phage-resistant)"				
	/note="Organ: testis; Vector: PCMV-SPOrt6; Site: 1; NotI;				
	Site: 2; SalI; Cloned unidirectionally; oligo-dT primed.				
	Average insert size 2.5 kb. Library enriched for				
	full-length clones and constructed by Life Technologies.				
	Note: This is a NIH_MGC Library."				
BASE COUNT	292 a 314 c 257 g 225 t				1 others
ORIGIN					
Alignment Scores:					
Pred. No.:	6.07e-102				1089
Score:	1111.50				229
Percent Similarity:	77.95%				Matches:
Best Local Similarity:	71.12%				Conservative: 22
Query Match:	42.95%				Mismatches: 46
DB:	14				Indels: 25
					Gaps: 4

QY	1	metalaserthr	il	elys	lualaleu	serVal	Val	Ser	glun	sp	lin	Ser	leu	Phe	glu	20							
Db	76	ATGGCAGACATATTAA	AGGAAGCCTTAT	CAGTTGTG	ATGAGGACCA	AGTCGTTGTTGAG	135																
QY	21	Cys	ala	tyr	gl	Ser	Pro	hi	leu	ala	lys	thr	glu	met	thr	ala	Ser	Ser	Ser	glu	40		
Db	136	TCTGCTTACGGA	AGACCGCACACCT	GGCTTACG	AGACAGAGAT	CACCGCTTCCTTCACGACGAC	195																
QY	41	Tyr	gl	Val	Thr	Ser	lys	met	Ser	Pro	ala	Val	Pro	glu	lna	sp	Pro	leu	Ser	glu	60		
Db	196	TATGGACACACTT	CCAAAGATGAGCC	ACCGCTCC	CTTCANACAGAGATT	GGCTGTCTCAACC	255																
QY	61	Pro	ala	arg	Val	Thr	ile	lys	met	glu	Cys	Asn	Pro	asn	Gln	Val	Asn	gl	Ser	arg	80		
Db	256	CCAGCAGGAGGT	CACCATCAAT	AAATGGAAT	TAAACCT	AGCCAGAGTGAAT	GGCTCAAGAAC	315															
QY	81	Ser	Pro	asp	Pasp	Cys	Ser	Val	ala	lys	gl	yl	lys	Met	Val	Ser	Ser	Ser	asp	asn	100		
Db	316	TCTCTGATGAAT	GTGACGTGTGGC	CAAAAGCGGGA	AGATGTTGGG	CGCCACCAACCGT	375																
QY	101	gl	Met	asn	Tyr	gl	Ser	tyr	Met	glu	glu	ly	lys	His	le	Pro	Pro	Pro	asn	Met	thr	120	
Db	376	GGGATGAATCT	ACGGAGCTTAC	ATGAGGAGGA	AGACACAT	ATCCACCCCAAACTG	ACACAG	435															
QY	121	Asn	gl	arg	arg	Val	ile	Val	Pro	ala	asp	Pro	thr	leu	Tyr	Ser	Thr	asp	His	Val	arg	140	
Db	436	AACGAGCGAG	AGTTATCGT	GGCAGACAG	ATCTTAC	CGCTATGGAG	ATGACACCAT	GTGGG	495														
QY	141	glu	Trp	leu	glu	Trp	ala	Val	lys	glu	tyr	gl	yl	leu	Pro	asp	Val	asp	lle	leu	leu	160	
Db	496	CAGTGGCTGG	AGATGGCGGT	GTGAAGAA	ATATGCGCTTC	CCAGCGTCAAC	ATCTTGTAATTC	555															
QY	161	Gln	asn	lle	asp	gl	yl	ys	glu	leu	Cys	lys	met	thr	lys	asp	asp	phe	lin	arg	leu	180	
Db	556	CAGAACATTC	ATCGAAGAACT	GTGCAGAGAT	AGCCAACGAC	AGACTTCCAGAGGCT	CACG	615															
QY	181	-	Pro	Ser	tyr	Asn	ala	asp	lle	leu	leu	Ser	His	leu	His	tyr	leu	Arg	glu	Thr	Pro	200	
Db	616	CCCCAGCTAC	ATGAGCGGC	ACATCTCT	CTGCACAT	CTCCAC	TACTCCACTAC	CTCAGAGAACT	CTCT	675													
QY	200	u	Pro	His	leu	Thr	Ser	asp	Pasp	-	Val	asp	lys	ala	leu	lna	sp	Ser	Pro	arg	leu	met	220
Db	676	TCCACATTTG	ATCACTAC	ATGATGT	GTGATAAAG	CCCTTCCAAAC	ACTCTCACCGGATATGC	735															
QY	220	is	-	ala	arg	asn	Thr	gl	yl	Val	ala	Thr	-	phe	le	phe	Pro	asn	Thr	-	Ser	Val	237
Db	736	ATGGCTAA	AACCCAGG	GGGGGGG	CGCCACGCTTT	ATTTATTTCCCAAT	TTCATTCGACAAATATCT	795															
QY	237	r	Pro	glu	lna	Thr	gln																

SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 915)
AUTHORS NIH-MGC http://mhc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LLM10429 row: m column: 13
High quality sequence stop: 820.
Location/Qualifiers
1..915
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4524852"
/clone_id="NIH_MGC_92"
/tissue_type="embryonal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: testis; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally; Oligo-dT primed. Average insert size 2.5 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC library."
BASE COUNT 246 a 252 c 227 g 190 t
ORIGIN
Alignment Scores:
Pred. No.: 2,27e-99 Length: 915
Score: 1085.00 Matches: 238
Percent Similarity: 77.26% Conserved: 10
Best Local Similarity: 74.14% Mismatches: 26
Query Match: 41.92% Indels: 51
DB: 12 Caps: 4
US-09-902-772-4 (1-478) x BG390291 (1-915)
QY 1 MetAlaSerThrIleLysGluAlaLeuSerValSerGluAspGlnSerLeuPheGlu 20
DB 59 ATGGCCAGCACTATTAGAGAGCCTTATCACTGTGAGTGAAGACCAAGTCCTGTGAG 118
QY 21 CysAlaIleTyrGlySerProHisLeuAlaLysThrGluMetThrAlaSerSerSerGlu 40
DB 119 TGTCCCTACGGAGACGACACCTGCTAAAGACAGAGATGACCGGCTCTCCAGCGAC 178
QY 41 TTrGlyGlnThrSerLysMetSerProArgValProGlnGlnAspTyrPleuSerGlnPro 60
DB 179 TATGACAGACTTCAGAGATAGGCCACGCTCCCTCAGCAGATGTGCTTCTCAACCC 238
QY 61 ProAlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsn 80
DB 239 CCAGCCAGGGGACCATCAAAATGGAATGTAACTAGCCAGTGAATGGCTCAAGAAC 298
QY 81 SerProAspAspCysSerValAlaIleGlyGlyLysMetValSerSerSerAspAsnVal 100
DB 299 TCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 358
QY 101 GlnMetAsnTyrGlySerTyrMetGluGluLysHisIleProProAsnMetThrThr 120
DB 359 GGGATGAACATGACGACCTACATGAGAGAGACACATGCCACC-CCAAACATGACCGC 417
QY 121 AsnGluArgArgValIleValProAla-AspProThrLeuTyrSerThrAspHisValar 140
DB 418 AACGAGCGCAGAGTTATGTCGCCAGCAAGATCTACCTATGAGTACAGACCATGTGCG 477

QY 140 GlnTrpLeuGluTrpAlaValLysGluTyrGlyLeuProAspValAspIleLeuLeuPh 160
DB 478 GGAGTGGCTGGAGTGGGGGTAAAGATATGCGCTTCAGACGTCACATCTTGTAT 537
QY 160 eGlnAsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArgLeuTh 180
DB 538 CCAGAACATGATGGAGGAAGAACTGTCAAGATGACCAAGAGACGACTTCCAGAGCTCAG 597
QY 180 rProSerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGluThrPro-L 200
DB 598 CCCCAGCTACATGCGGCAGAT-CTTCTCATCATCTCCACTCCTCAGAGACATCTTC 656
QY 200 euProHisLeuThrSerAspAspValAspLysAlaLeuGlnAsnSerProAlaGluMetH 220
DB 657 TTCCACATTTGACTTCATATGATGTATTAAGCTTACAAAACCTCCACGCTTAATGC 716
QY 220 lSAlaArgAsnThrGlyAlaThrPheLeuPheProAsnThrSerValTyrProGluA 240
DB 717 ATGCTAGAAACACA----- 730
QY 240 lArhGlnArgIleThrThrArgProAspLeuProTyrGluAlaArgArgSerAlaT 260
DB 731 -----GATTACA-TATGAGCCCCCAGAGATCAAGGC 763
QY 260 rPThrSerHisSerHisProThrGlnSerLysAlaIleThrGlnProSerSerThrValP 280
DB 764 TGGACGCTACAGG-CACCCACGCCAGCTCAAGCTGTACATCTTCCACAGTCCCAA 822
QY 280 rOlyThrGlu-----AspGlnArgProG 288
DB 823 TGAAGACAGCTCTCGTATATCATATCTTGGCACAGTCCCTCGCAATCGGGGCCCA 882
QY 288 lNleuAspProTyrGlnIleLeuGlyProThrSerSerArgLeuAlaAsnProGlySer 307
DB 883 CTTTGATGCC-----GGCCGGGGAGC-----CCGCGCTCC 914
RESULT 6
BG388025 715 bp mRNA linear EST 12-MAR-2001
LOCUS 602412867F1 NIH_MGC_92 Homo sapiens CDNA clone IMAGE:452123 5',
DEFINITION mRNA sequence.
ACCESSION BG388025
VERSION BG388025.1 GI:13281471
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 715)
AUTHORS NIH-MGC http://mhc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LLM10420 row: f column: 18
High quality sequence stop: 693.
Location/Qualifiers
1..715
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4521233"
/clone_id="NIH_MGC_92"
/tissue_type="embryonal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: testis; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally; Oligo-dT primed.

Average insert size 2.5 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: This is a NIH-MGC Library."

BASE COUNT 206 a 188 c 172 g 149 t

Alignment Scores:

Pred. No.:	2,74e-96	Length:	715
Score:	1053.00	Matches:	195
Percent Similarity:	97.58%	Conservative:	7
Best Local Similarity:	94.20%	Mismatches:	5
Query Match:	40.69%	Indels:	0
DB:	12	Gaps:	0

US-09-902-772-4 (1-478) x BG388025 (1-715)

```

QY 1 MetAlaSerThrIleYsgIuaIaLeuSerValaSerGIuaSpGlnSerLeuPheGlu 20
DB 94 ATGGCAGGACCTTAAAGAGCCCTTATCAGTGTGAGTAGAGACCAAGCTGCTTTGAG 153
QY 21 CysAlaTyGlySerProHisLeuAlaLysThrGluMetThrAlaSerSerSerGlu 40
DB 154 TGTGCTTACGGAAGCCACACCTGCTTAAGACAGATGACCGCTCTCCACAGGAC 213
QY 41 TyrGlyGlnThrSerIysMetSerProArgValProGlnGlnAspThrLeuSerGlnPro 60
DB 214 TATGACAGACTTCCAAAGATGAGCCACGCGCTCTCAGAGGATGGCTCTCAACC 273
QY 61 ProAlaArgValThrIleYsgIuaIaLeuSerValaSerProGlnGlnAlaSerArgAsn 80
DB 274 CCGACGAGGCTCCACATCAAAATGGAATGTAACCTTACAGCAGTGAATGGCTTAAGGAC 333
QY 81 SerProAspAspCysSerValaLysGlyLysMetValSerSerSerAspAsnVal 100
DB 334 TCTCCGATGATGATGACAGTGTGGCCAAAGCGGGAAGATGTGGGACAGACACCGGT 393
QY 101 GlyMetAsnTyGlySerIysMetGluGluLysHisIleProProAsnMetThrThr 120
DB 394 GGGATGACATGAGCGGACATGAGGAGAGACACATGACATGACATGACATGACATGAC 453
QY 121 AsnGluArgValIleValProAlaAspProThrLeuTrpSerThrAspHisValArg 140
DB 454 AACGACGAGAGAGTATTCGTGGCAGACATGCTTACGCTTACGCTTACGCTTACGCTTAC 513
QY 141 GlnTrpLeuGluTrpAlaValLysGlyLysLeuProAspValAspIleLeuLeuPhe 160
DB 514 CAGTGGCTGAGAGTGGCGGTGAAGATATGAGCTTCCAGACATCAATCTTATATTC 573
QY 161 GlnAsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArgLeuThr 180
DB 574 CAGACATGATGAGGAGAGAGTGTCAAGATGACCAAGAGACATTCACAGAGGCTCAC 633
QY 181 ProSerTyAsnAlaAspIleLeuLeuSerHisLeuHisTyLeuArgGluThrProLeu 200
DB 634 CCGAGTCAACAAGCCGACATCCCTCTCTCACAATCTCCATTAACCTCAGAGACTCTCTT 693
QY 201 ProHisLeuThrSerAspAsp 207
DB 694 CCACATATGACTTCAGATGAT 714

RESULT 7
LOCUS AW948986 629 bp mRNA linear EST 31-MAY-2000
DEFINITION QV4-FT0005-110500-201-102 FT0005 Homo sapiens cDNA, mRNA sequence.
ACCESSION AW948986
VERSION AW948986.1 GI:8126760
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 629)
AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,

```

TITLE
JOURNAL
MEDLINE
COMMENT

Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?pl=ft2-QV4-FT0005-110500-201-102&ts=2000-05-11&tl=1)
Seq primer: puc 18 forward
High quality sequence stop: 628.

FEATURES

source
1..629
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="FT0005"
/dev_stage="Adult"
/note="Organ: prostate; tumor: Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT 166 a 173 c 171 g 119 t
ORIGIN

Alignment Scores:

Pred. No.:	5.15e-92	Length:	629
Score:	1010.00	Matches:	190
Percent Similarity:	96.57%	Conservative:	7
Best Local Similarity:	93.14%	Mismatches:	1
Query Match:	39.03%	Indels:	0
DB:	10	Gaps:	0

US-09-902-772-4 (1-478) x AW948986 (1-629)

```

QY 6 LysGluAlaLeuSerValaSerGIuaSpGlnSerLeuPheGlnCysAlaTyGlySer 25
DB 19 CAGGAAGCCTTATCACTTGAGTAGAGACCAAGTGTGTTGAGTGGCTACGGAAG 78
QY 26 ProHisLeuAlaLysThrGluMetThrAlaSerSerSerSerGluTyGlnThrSer 45
DB 79 CCGACCTGGCTTAAGACAGATGAGCGGCTCTCTCCAGCAGCATATGAGACACTTCC 138
QY 46 LysMetSerProArgValProGlnGlnAspTrpLeuSerGlnProProAlaArgValThr 65
DB 139 AACATAGCCACAGCGTCCCTCAGCAGAGTGGCTTCTCAACCCACAGGCTCAC 198
QY 66 IleLysMetGluCysAsnProAsnGlnValaSerArgAsnSerProAspAspCys 85
DB 199 ATCAAAATGGAATGTAACCTTACGAGTGAATGCTCAAGAGACTCTCTGATGATGC 258
QY 86 SerValaLysGlyLysLysMetValaSerSerSerAspAsnValaLysLysTyGly 105
DB 259 AGTGTGGCCAAAGCGGGAAGATGTGGGACAGCCACAGACCGTGTGGATGAACTACGCG 318
QY 106 SerTyMetGluGluLysHisIleProProProAsnMetThrThrAsnGluArgVal 125
DB 319 AGCTACATGAGAGAGAGACATGCCACCCCAACATATGACCAAGAGGCGCAGATT 378

```

```

QY      126  lilevalproliaaspprothleuttrserthrasphisvalarglntprleuglutr 145
Db      379  ATCGTGCAGACGATCTACACCTATGAGTACAGACCATGTGCGGACATGCGTGGAGTGG 438
QY      146  AlAlaValysglutYrglYleuproaspvalaspilleuleupheglinaasply 165
Db      439  GCGGGAAGAGATATGGCTTCACAGCTCAACATCTGTATTCACAGACATCATGAGG 498
QY      166  LysgluleucYsLysMetThrLysAspAspPheglInargleuthrProSerTyrAsnAla 185
Db      499  AAGGAAGTGCAGAGATGACCAATGAGAGACATTCAGAGAGTCCACCCAGCTACAGAGCC 558
QY      186  AspileuleuSerHisLeuHisTyrleuarglutrProleuProHisleuThSer 205
Db      559  GACATCCCTTCTCTCACAATCTCCACTCACTCACTCACTCCCTTCCACATTTGACTCA- 617
QY      206  AspaSpValasp 209
Db      618  GATGATGGTGAT 629

RESULT 8
LOCUS      BG259459                928 bp    mRNA    linear    EST 13-FEB-2001
DEFINITION 602378556F1 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:4509541 5',
            mRNA sequence.
ACCESSION  BG259459
VERSION    BG259459.1  GI:12769275
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE  1 (bases 1 to 928)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgabs-remail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLU)
            DNA Sequencing by: Incyte Genomics, Inc.
            Found through the I.M.A.G.E. Consortium/LNLU at:
            http://image.lnl.gov
            Plate: L1AM10389 row: 0 column: 14
            High quality sequence stop: 717.
FEATURES   location/Qualifiers
            source
            1..928
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:4509541"
            /clone_lid="NIH_MGC_92"
            /tissue_type="embryonal carcinoma, cell line"
            /lab_host="DH10B (phage-resistant)"
            /note="Organ: testis; Vector: pCMV-SPORT6; Site_1: NotI;
            Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
            Average insert size 2.5 kb. Library enriched for
            full-length clones and constructed by Life Technologies.
            Note: this is a NIH_MGC Library."
BASE COUNT 265 a 242 c 220 g 201 t
ORIGIN
Alignment Scores:
Pred. No.: 5,51e-90 Length: 928
Score: 992.50 Matches: 221
Percent Similarity: 83.63% Conservativeness: 14
Best Local Similarity: 78.65% Mismatches: 34
Query Match: 38.35% Indels: 17
DB: 12 Gaps: 4
US-09-902-772-4 (1-478) x BG259459 (1-928)

```

```

QY      1  MetAlaSerThrIleValgluAlaLeuSerValValSerGluAspGlnSerLeuPheGlu 20
Db      88  ATGGCCACACACTATTAAAGAAAGCCCTTATACAGTTGTGAGTGAAGACCATCGTG-TTTGAAG 146
QY      21  CysAlaTyrGlySerProHisLeuAlaYsrThrGluMetThrAlaSerSerSerGlu 40
Db      147  TGTGCTTACGGAAGCCACACCTGGCTAAGACAGAGATGACCGGCTCTCCACAGCAGC 206
QY      41  TyrGlyGlnThrSerLysMetSerProArgValProGlnGlnAspThrLeuSerGlnPro 60
Db      207  TATGACACAGACTTCCAAATATGAGCCACGCGCTCCACAGAGATGGCTGTCTCAACCC 266
QY      61  ProAlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsn 80
Db      267  CCAGCAGAGCTC-ACCATCAAAATGAATGTAACCTTACCAGAGTGAATGGCTCAAGAAC 325
QY      81  SerProAspAspCysSerValAlaLysGlyLysMetValSerSerSerAspAsnVal 100
Db      326  TCTCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 385
QY      101  GlyMetAsnTyrGlySerTyrMetGluGluLysHisIleProProProAsnMetThrThr 120
Db      386  GGGATGACACTACGCGCAGCTACATGAGAGAGAACACATGCCACCCCAACATGACACAG 445
QY      121  AsnGluArgArgValIleValProAlaAspProThrLeuTyrSerThrAspHisValArg 140
Db      446  AACGAGCCAGAGTATATGTCGCCACAGATCTTACGCTATGAGATACAGACCATGTGGCG 505
QY      141  GlnTrpLeuGluTrpAlaValLysGluTyrGlyLeuProAspValAspIleLeuLeuPhe 160
Db      506  CAGTGGCTGAGTGGGCGGTGACAGAAATATGGCTTCCAGAGCTCAACATCTTGTATATC 565
QY      161  GlnAsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArgLeuThr 180
Db      566  CAGAACATCATGCGGAAAGAACTGTGCAAGATGACCAAGACAGACTTCCAGAGGCTCAC 625
QY      181  ProSerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGluThrProLeu 200
Db      626  CCAG--TACAATGGCCGACATCCTCTCTCACATCTCCACTACCTCAGAGAAATCTCTCT 683
QY      201  ProHis-LeuThrSerAspAspValAspLysAlaLeuGlnAsnSerProArgLeuMetHis 220
Db      684  GCACATTTTGACTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 742
QY      220  AlaArgAsnThrGlyGlyAlaThrPheIlePheProAsnThrSerValTyrProGluAla 240
Db      743  TGCCTTGA---ACCCAGAGGGGTGCTGTTTA-TTTCCAATATTAATTAAT---TTCGAAGC 795
QY      240  aThrGlnArgIleThrThrArgProAspLeuProTyrGlnGlnAlaArgSerAlaTr 260
Db      796  TACGCAACGATTA-----CCACTTGGCAGTGCAGAAAACCCGCGTGCATCTGTCG 843
QY      260  pThrSerHisSerHisProThrGlnSerLysAlaThrGlnProSerSerSerThrValPr 280
Db      844  GCAGAAC-----CAAGTCGCGGATGCGGCCAATCAGAAACAAAGTGC 888
QY      280  o 280
Db      889  C 889

RESULT 9
LOCUS      AUI36709                701 bp    mRNA    linear    EST 02-AUG-2002
DEFINITION AUI36709 PLACE1 Homo sapiens cDNA clone PLACE1004911 5', mRNA
            sequence.
ACCESSION  AUI36709
VERSION    AUI36709.1  GI:10997248
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE  1 (bases 1 to 701)

```

AUTHORS Ota, T., Nishikawa, T., Suzuki, Y., Ishii, S., Saito, K., Kawai, Y., Yamamoto, J., Wakamatsu, A., Nakamura, Y., Nagai, T., Sugano, S. and Isogai, T.

TITLE HRI human cDNA project

JOURNAL Unpublished (2000)

COMMENT Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3975
Fax: 81-438-52-3986
Email: genomics@hri.co.jp
HRI human cDNA project, 5'- & 3'-end one pass sequencing: Helix Research Institute; cDNA library construction: Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute.

FEATURES
source 1. 701
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1fb="PLACE1004911"
/clone_1lb="PLACE1"
/tissue_type="Placenta"
/note="Vector: pME18SFL3"

BASE COUNT 199 a 185 c 166 g 145 t 6 others

ORIGIN

Alignment Scores:
Pred. No.: 1,03e-89 Length: 701
Score: 988.00 Matches: 189
Percent Similarity: 94.20% Conservative: 6
Best Local Similarity: 91.30% Mismatches: 10
Query Match: 38.18% Indels: 2
DB: Gaps: 0

US-09-902-772-4 (1-478) x AU136709 (1-701)

QY 1 MetAlaSerThrIleGluAlaLeuSerValSerGluAspGlnSerLeuPheGlu 20
|||||
DB 85 ATGGCCAGACCTATTAAAGAACCTTATCAGTTGAGTGAAGACCAAGCGTTGTTGAG 144
|||||
QY 21 CysAlaTyrGlySerProHisLeuAlaLysThrGluMetThrAlaSerSerSerGlu 40
|||||
DB 145 TGTGGCTAGGAAAGCCACACCTGGCTAAGACAGACATACCCGCTCTCTCCACGAC 204
|||||
QY 41 TyrGlyGlnThrSerLysMetSerProArgValProGlnGlnAspTyrLeuSerGlnPro 60
|||||
DB 205 TATGGACAGACTCCAAAGATGAGCCACGCGTCCCTCAGCAGAGATTGGCTGCTCAACC 264
|||||
QY 61 ProAlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsn 80
|||||
DB 265 CCAGCCAGGAGTACCATCAAAATGGAATGAACCCAGGAGTGAATGGCTCAAGGAAC 324
|||||
QY 81 SerProAspAspCysSerValAlaLysGlyLysMetValSerSerSerAspAsnVal 100
|||||
DB 325 TCTCCGATGAAATGCGAGTGTGGCAAAGCGGGAAGATGTGGCGACCCACAGACCGTT 384
|||||
QY 101 GlyMetAspTyrGlySerTyrMetGluGluLysHisIleProProAsnMetThrThr 120
|||||
DB 385 GGGATGAATACGCGACGCTACATGAGGAGAGACATGCCACCCCAACATGACACGAC 444
|||||
QY 121 AsnGluArgArgValIleValProAlaAspProThrLeuTyrSerThrAspHisValArg 140
|||||
DB 445 AACGAGCGGAGATTATCGTGCACAGATCTAGCGTATGAGATGACACATGTGCGG 504
|||||
QY 141 GlnTyrLeuGluTyrPalValAlaLysGluTyrGlyLeuProAspValAspIleLeuLeuPhe 160
|||||
DB 505 CAGTGGCTGAGTGGCGGTGAAGAATATGCGCTTCCAGACATCAACATCTTGTATTTC 564
|||||
QY 161 GlnAsnIleAspGlyLysGluLeuCysLysMetThr-LysAspAspPheGlnArgLeuThr 180
|||||
DB 565 CAGAACATCGATGAGGAANCACTGTGCACAGATGACCAAGGACGACTTTCNAGANGCTCAA 624
|||||

QY 180 rProSerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeu-ArgGluThrPro 200
|||||
DB 625 CCCAGCTTACAGCCGACATCTTCTCTCATCTCCATCTCACTGACAGAGACTCTC 684
|||||
QY 200 eufProHisLeuThrSer 205
|||
DB 685 TTTCACATTTGNCNTCA 701
|||||

RESULT 10
BI713036 642 bp mRNA linear EST 11-MAR-2002
BI713036
LOCUS
DEFINITION
le01d08.y1 HR85 1slet Homo sapiens cDNA clone IMAGE:5085758 5' similar to SW:ERG_HUMAN P11308 TRANSCRIPTIONAL REGULATOR ERG ; mRNA sequence.
BI713036
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
Melton, D., Brown, J., Kently, G., Permutt, A., Lee, C., Kaestner, K., Lemishka, I., Scaerce, M., Brestelli, J., Gradwohl, G., Clifton, S., Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, J., Blistain, A., Schmitt, A., Treising, B., Ritter, E., Ronko, I., Bennett, J., Cardenas, M., Gibbons, M., McCann, R., Cole, R., Tsagaris, R., Williams, T., Jackson, Y., and Bowers, Y.
Endocrine Pancreas Consortium
Unpublished (2000)
Other: le01d08.x1
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biohp.harvard.edu
Library was constructed by Dr. Hiroshi Inoue DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: Dr. Hiroshi Inoue (hinoue@im.wustl.edu)
Seq primer: -40RP from Gibco
High quality sequence stop: 473.
Location/Qualifiers
1. 642
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1fb="HR85 1slet"
/tissue_type="Purified pancreatic 1slet"
/lab_host="DH10B"
/note="organ: Pancreas; Vector: pBluescript SK(-); Site: 1: Notr. Site: 2: XhoI; cDNA made by oligo-dT priming. Size-selected on agarose gel. Average insert size -1kb. 5' XhoI site was destroyed after directional cloning. Amplified once. Contact information: Hiroshi Inoue, MD, Metabolism Div. (Alan Permutt Lab), Washington University School of Medicine, Box 8127, 660 South Euclid Ave., St. Louis, MO 63110, E-mail: hinoue@im.wustl.edu, Tel: 314-362-1916, Fax: 314-747-2692."

BASE COUNT 178 a 171 c 163 g 129 t 1 others

ORIGIN

Alignment Scores:
Pred. No.: 2.3e-89 Length: 642
Score: 984.00 Matches: 183
Percent Similarity: 97.46% Conservative: 9
Best Local Similarity: 92.89% Mismatches: 5
Query Match: 38.02% Indels: 0
DB: Gaps: 13


```

Db 200 TATGGCCAGACATCCAGATGATGCCAGATCCCTCAGCAGACGTGCTGTCTCAAGCC 259
Oy 61 ProAlaArgValThrIleLeuMetGluCysAsnProAsnGlnValAsnGlySerArgAsn 80
Db 260 CCAGCCAGGGGTGCACATCAAGATGAGTGCACACCTGAGTGAATGGTTCACAGGAAAC 319
Oy 81 SerProAspAspCysSerValAlaLeuGlyGlyMetValSerSerSerAspAsnVal 100
Db 320 TCACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 379
Oy 101 GlyMetAsnTyrGlySerTyrMetGluGluLeuHisIleProProProAsnMetThrThr 120
Db 380 GGGATGAGTACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 439
Oy 121 AsnGluArgArgValIleValProAlaAspProThrLeuTyrSerThrAspHisValArg 140
Db 440 AATGACGCGCAGAGTGTATCCCTCCGACAGATCTACTCTGTGGAGCAGACACATGTCCGA 499
Oy 141 GlnTyrLeuGlnTyrAlaValAlaLeuGlyTyrGlyLeuProAspValAspIleLeuPhe 160
Db 500 CAGTGGCTGGAGTGGCGGGGGAAGAATATGGCTCTGCTGATGTGACGCTTACTATTTT 559
Oy 161 GlnAsnIleAspGlyLeuGluLeuCysLysMetThrLysAspAspPheGlnArgLeuThr 180
Db 560 CAGAAATATGATGATGGAAGAGCTGTGCAAGATGACAAAGATGACTTCCACGGCGCTCAGC 619
Oy 181 ProSerTyrAsnAlaAspIleLeuSerHisLeuHisTyrLeuArg 196
Db 620 CCGAGCTACAAATGCCAGCATTTCTTCTCTACATCTTCACACTTACTCTCAGA 667

RESULT 12
BM088063 592 bp mRNA linear EST 19-NOV-2001
DEFINITION 50136 MARC 2B0V Bos taurus cDNA 5', mRNA sequence.
ACCESSION BM088063
VERSION BM088063.1 GI:16998691
KEYWORDS EST.
SOURCE cow.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 592)
Smith,T.P.L., Grose,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
Caess,E., Wray,J.B., White,J., Cho,J., Fahrenkrug,S.C., Bennett
,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
Perteira,G., Holt,I., Karameycheva,S., Liang,F., Quackenbush,J. and
Keele,J.W.
Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
Genome Res. 11 (4), 626-630 (2001)
21180013
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel.: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt-trimmed with phred
v0.980904.e. Vector identified by cross_match with the -mnscore 18
and -minmatch 12 options.
PCR primers
FORWARD: AGGAACAAGCTATGACCAAT
BACKWARD: GTTTCCTCAGTACGACGACG
Plate: 139 row: D column: 21
Seq primer: ATTAGTGACACTATAG.
FEATURES
Source
1..592
location/Qualifiers
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 2B0V"
/tissue_type="pooled"
/lab_host="DH10B"

```

```

BASE COUNT 152 a 177 c 159 g 104 t
ORIGIN
Alignment Scores:
Pred. No.: 5,17e-89 Length: 592
Score: 980.00 Matches: 181
Percent Similarity: 96.89% Conservative: 6
Best Local Similarity: 93.78% Mismatches: 6
Query Match: 37.87% Indels: 0
DB: 13 Gaps: 0
US-09-902-772-4 (1-478) x BM088063 (1-592)
Oy 1 MetaSerThrIleLeuGluAlaLeuSerValIleSerGluAspGlnSerLeuPheGlu 20
Db 12 ATGGCCAGCAGCTATTAAGGAAGCGTTATCAGTGTGAGGACGACGACGACGACGACG 71
Oy 21 CysAlaTyrGlySerProHisLeuAlaLeuThrGluMetThrAlaSerSerSerGlu 40
Db 72 TGGCGCTACGGCAGCGCGCAGCTGCCCAACGCGACATGACCCGCTCTCTCCGGCGAC 131
Oy 41 TyrGlyGlnThrSerLysMetSerProArgValProGlnGlnAspTyrLeuSerGlnPro 60
Db 132 TATGACAGACATCAAGATGAGCCGCGGCTCCCTCAGAGAGAGTGGGTCTCAAGCC 191
Oy 61 ProAlaArgValThrIleLeuMetGluCysAsnProAsnGlnValAsnGlySerArgAsn 80
Db 192 CCGAGCGGTCACCATCAAGATGAGATGTAACCTTAACCAAGTGAACGCTTCAAGGAAC 251
Oy 81 SerProAspAspCysSerValAlaLeuGlyGlyLysMetValSerSerSerAspAsnVal 100
Db 252 TCCCGCTGATGATGACGACATGCGCAAGGTGGGAATGTGGGCGACCCGCGACCGCTC 311
Oy 101 GlyMetAsnTyrGlySerTyrMetGluGluLysHisIleProProProAsnMetThrThr 120
Db 312 GGGATGATACAGCAGCATCATGAGAGAGAAACATGACATGCCCAACATGACACACC 371
Oy 121 AsnGluArgArgValIleValProAlaAspProThrLeuTyrSerThrAspHisValArg 140
Db 372 AAGGACGCGCGGCTCATCTGCTCCGACATCCACGCTTATGAGATGACATGTCGGC 431
Oy 141 GlnTyrLeuGlnTyrAlaValAlaLeuGlyTyrGlyLeuProAspValAspIleLeuPhe 160
Db 432 CAGTGGCTGAGTGGGACATCAAGAATACGGGCTTCCGAGCTTGACATTTATTATTC 491
Oy 161 GlnAsnIleAspGlyLeuGluLeuCysLysMetThrLysAspAspPheGlnArgLeuThr 180
Db 492 CAGAAATATGAGGAGGAGAGCTGTGCAAGATGACCAAGATGACTTCCAGAGGCTCACC 551
Oy 181 ProSerTyrAsnAlaAspIleLeuSerHisLeuHis 193
Db 552 CCGAGCTACAAAGCTGACATCTTGTGTGCACTCCAC 590

RESULT 13
AJ456498 827 bp mRNA linear EST 22-APR-2002
DEFINITION AJ456498 riken1 Gallus gallus cDNA clone Bml3f2, mRNA sequence.
ACCESSION AJ456498
VERSION AJ456498.1 GI:20266594
KEYWORDS EST.
SOURCE chicken.
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianidae; Gallus.
1 (bases 1 to 827)
REFERENCE Buerstedde,J.M.
AUTHORS Gallus gallus bursal lymphocyte EST
TITLE Gallus gallus bursal lymphocyte EST
JOURNAL Unpublished (2002)

```

COMMENT Contact: Buerstedde JM
Cellular Immunology
Heinrich-Pette-Institute
Martinistr. 52, 20251 Hamburg, Germany
Email: URL: <http://genetics.hpi.uni-hamburg.de/dt40est.html>.
Location/Qualifiers

FEATURES
Source
1. 827
/organism="Gallus gallus"
/db_xref="taxon:9031"
/clone_lib="8ml3r2"
/clone_lib="riken1"
/cell_type="Bursal lymphocyte"
/dev_stage="2-3 weeks old"
/note="CB inbred strain"

BASE COUNT 233 a 236 c 177 g 181 t

ORIGIN

Alignment Scores:
Pred. No.: 1,41e-87 Length: 827
Score: 968.00 Matches: 194
Percent Similarity: 62.87% Conservative: 21
Best Local Similarity: 56.73% Mismatches: 39
Query Match: 37.40% Indels: 88
DB: Gaps: 5

US-09-902-772-4 (1-478) x AJ456498 (1-827)

QY 138 HisValArgGlnTrpLeuGluTrpAlaValLysGluTrpGlyLeuProAspValAspIle 157
|||||
DB 5 CAGCTGGCCAGAGTGGCTGGAATGGCCATTAAGAGATGATTAATGAGATTCAGACACC 64
|||||
QY 158 LeuLeuPheGlnAsnIleAspGlyLysGluLeuGlyLysMetThrLysAspAspPheGln 177
|||||
DB 65 ACCCTCTCCAGATATGATGATGAGCAAGAGCTCTGCAAAAGCAAGATGATGATCTCTC 124
|||||
QY 178 ArgLeuThrProSerThrAsnAlaAspIleLeuLeuSerHisLeuHisThrLysLeuArgGlu 197
|||||
DB 125 CGAAGCCAGCTCCCTACCAACAGAGAGTCTGTCTCACCCTCAGCTCCTCAGGAA 184
|||||
QY 198 ThrProLeuProHisLeuThrSerAspValAspLysAlaLeuGlnAsnSerProArg 217
|||||
DB 185 AGT----- 187
|||||
QY 218 LeuMetHisAlaArgAsnThrGlyAlaThrPheIlePheProAsnThrSerValTyr 237
|||||
DB 188 -----AGCTCACTGCTGCTACATCACTCCTCCTCCAC 220
|||||
QY 238 ProGluAlaThrGlnArgIleThrThrArgProAspLeuProTyrGluGlnAlaArgArg 257
|||||
DB 221 ACAGAAAGCTTCACCTCTGCGCACCAAGAGAGTCTCTCTGT----- 265
|||||
QY 258 SerAlaThrPheSerHisSerHisProThrGlnSerLysAlaThrGlnProSerSerSer 277
|||||
DB 266 -----GCAGGACACCAAAACGTCAGCAAG 289
|||||
QY 278 ThrValProLysThrGluAspGlnArgProGlnLeuAspProTyrGlnIleLeuGlyPro 297
|||||
DB 290 ACA-----ACAGAAACGCAACGCGCTCAGCAACATCCATCAATCTCGGGGCC 340
|||||
QY 298 ThrSerSerArgLeuAlaAsnProGlySerGlyGlnIleGlnLeuThrGlnPheLeuLeu 317
|||||
DB 341 AGTAGTAGTGTCTGCGCAATCTGGAGTGGGCGAGTACATCACTGAGTCTCTCTC 400
|||||
QY 318 GlnLeuLeuSerAspSerSerAsnSerAsnGlyIleThrTrpGluGlyThrAsnGlyGlu 337
|||||
DB 401 GAGTGTGTGCGGAGCTTCCAGCGCAGCTGATCATCGGAGGAGGACCATGGGGA 460
|||||
QY 338 PheLysMetThrAspProAspGluValAlaArgArgTrpGlyLysSerLysPro 357
|||||
DB 461 TTCAGATGTGACAGACCGAGTGAAGTGGCAGCGCTGGGAGAACCAACGCAAGCCC 520
|||||
QY 358 AsnMetAsnThrAspLysLeuSerArgAlaLeuArgTyrTyrTyrAspLysAsnIleMet 377
|||||

DB 521 AACATGAATTATGACAAAGCTGAGCCAGCCCTTCGATACACTATGATGATAGACATTATG 580
|||||
QY 378 ThrLysValHisProGluSerSerMetLysTyrProSerAspLeuProTyrMet 397
|||||
DB 581 ACCAAAGTG----- 589
|||||
QY 398 SerSerThrLysGlyLysArgTyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAla 417
|||||
DB 590 -----CATGCAAAAGATGATGCTTACAAATTTGACTTTCATGGCATGGCCAGGCT 640
|||||
QY 418 LeuGlnPro----- 420
|||||
DB 641 CTCAGGCTCATCCCATCCATGATCATATGATGATGATGATGATGATGATGATGATGATG 700
|||||
QY 421 -----HisAlaHisProGlnLysMetAsnPheValAlaProHisProProAlaLeu 437
|||||
DB 701 CCTTCTTACCATGCTCCACAGCAGAGAGTGAATGATGATGATGATGATGATGATGATGATG 760
|||||
QY 438 ProValThrSerSerSerPheAlaAlaProAsnProTyrTrpAsnSerProThrGly 457
|||||
DB 761 CCTGTACATCATCATCATCTTCTCGAGAGACCTCCTATGATGATGATGATGATGATGATGATG 820
|||||
QY 458 GlyIle 459
|||||
DB 821 AGCAAT 826
|||||

RESULT 14
BB639043 675 bp mRNA linear EST 31-AUG-2001
LOCUS BB639043 RIKEN full-length enriched, 3 days neonate thymus Mus
DEFINITION musculus cDNA clone A630068021 5', mRNA sequence.
ACCESSION BB639043
VERSION BB639043.1 GI:15401323
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 675)
AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,
Hiramoto,K., Horii,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,
M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,
D., Shibata,K., Shingawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
unpublished (2001)
CONTACT: Yoshinori Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-resgsc.riken.go.jp,
URL: <http://genome.gsc.riken.go.jp/>
Carninci,P., Shibata,Y., Hayata,Y., Sugahara,Y., Shibata,K., Itoh,
M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Matsubara,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsunaga,
S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multiplexed sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,
Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a

Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.

FEATURES
source

location/Qualifiers

1. 629
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="6330501C04"
/clone_1lb="RIKEN full-length enriched, adult male medulla oblongata"
/sex="male"
/tissue_type="medulla oblongata"
/dev_stage="adult"
/note="Site 1: SalI; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5'
GAGAGAGAGAGATCCACAGCTCTTTTCTTTTCTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATCTCGAGTTATTAATTAATCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified Bluescript KS(+) after bulk excision from Lambda FLX I. Cloning sites, 5' end: SalI; 3' end: BamHI"

BASE COUNT 172 a 150 c 171 g 135 t 1 others
ORIGIN

Alignment Scores:

Pred. No.: 4,16e-84 Length: 629
Score: 932.00 Matches: 174
Percent Similarity: 95.77% Conservative: 7
Best local Similarity: 92.06% Mismatches: 8
Query Match: 36.01% Indels: 0
DB: 10 Gaps: 0

US-09-902-772-4 (1-478) x BB577887 (1-629)

QY 1 MetAlaSerThrIleuysGluAlaLeuSerValSerGluAspGlnSerLeuPheGlu 20
DB 61 ATGGCCAGCACTATTAAAGGAGGCGCTTGTCAGTGTGAGCGAGCACGATCACTATTGAG 120
QY 21 CysAlaTyrGlySerProHisLeuAlaLysThrGluMetThrAlaSerSerSerGlu 40
DB 121 TGTGCTTACGGAACGACCACTGGCTTAAGACAGATGACCGCATCTCTTCCAGTGAC 180
QY 41 TyrGlyGlnThrSerLysMetSerProArgValProGlnGlnAspTyrLeuSerGlnPro 60
DB 181 TATGGCCAGACATCCAGATGATGCCAGAGTCCCTCAGCAGAGTGGCTGTCTCAAGCC 240
QY 61 ProAlaArgValThrIleuysMetGluCysAsnProAsnGlnValAsnGlySerArgAsn 80
DB 241 CCAGCCAGGGGACACCAATGAGTGCACACCTAGTCAGTGAATGGTTCCAGGAAAC 300
QY 81 SerProAspAspCysSerValAlaLysGlyLysMetValSerSerSerAspAsnVal 100
DB 301 TCACCTGATGAGTGCATGTGTAACAAGGTGGAGATGGTGGCAGCCCGGATACTGTG 360
QY 101 GlyMetAsnTyrGlySerTyrMetGluLysHisIleProProAsnMetThrThr 120
DB 361 GGGATGAGCTACGCGACCTACATGAGAGAGAAAGCATGTGCGCCTCCCAATATGACACA 420
QY 121 AsnGluArgArgValIleValProAlaAspProThrLeuTyrSerThrAspHisValArg 140

DB 421 AATGAGCGCAGAGTGCATCTCCCTGACAGATCTACTGTGTGAGACAGACCAATGTCCGA 480
QY 141 GlnTrpLeuGlnTyrPAlaValLysGluTyrGlyLeuProAspValAspIleLeuPhe 160
DB 481 CAGTGGCTGGAGTGGCGCGGTGAAGAATATGGCCCTCCGATGTGGACGCTTACTATT 540
QY 161 GlnAsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArgLeuThr 180
DB 541 CAGATATCGATGGAGAGAGACAGTGCMAAGATGACATFANGATGACTTCACGCGCTCACG 600
QY 181 ProSerTyrAsnAlaAspIleLeuLeu 189
DB 601 CCGAGCTACAAATGCCGACATTCCTCTC 627

Search completed: November 9, 2002, 19:46:50
Job time : 1601.56 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 9, 2002, 12:30:11 ; Search time 2817 Seconds

(without alignments)
185,960 Million cell updates/sec

Title: US-09-902-772-1_COPY_645_662

Perfect score: 18

Sequence: 1 ctcagagagagagagacc 18

Scoring table: IDENTITY_NDC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 segs, 14551402878 residues

Total number of hits satisfying chosen parameters: 995600

Minimum DB seq length: 0

Maximum DB seq length: 100

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:*

1: gb.ba:*

2: gb.htg:*

3: gb.in:*

4: gb.om:*

5: gb.ov:*

6: gb.pat:*

7: gb.ph:*

8: gb.pl:*

9: gb.pr:*

10: gb.ro:*

11: gb.sts:*

12: gb.sy:*

13: gb.un:*

14: gb.vl:*

15: em.ba:*

16: em.fun:*

17: em.hum:*

18: em.in:*

19: em.mu:*

20: em.om:*

21: em.or:*

22: em.ov:*

23: em.pat:*

24: em.ph:*

25: em.pl:*

26: em.ro:*

27: em.sts:*

28: em.un:*

29: em.vl:*

30: em.htg.hum:*

31: em.htg.inv:*

32: em.htg.other:*

33: em.htg.mus:*

34: em.htg.pin:*

35: em.htg.rtd:*

36: em.htg.mam:*

37: em.htg.vrt:*

38: em.sy:*

39: em.htgo.hum:*

40: em.htgo.mus:*

41: em.htgo.other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	14.8	82.2	91	14	AF218944
2	13.8	76.7	35	11	C75748
3	13.8	76.7	78	10	MM0130793
4	13.4	74.4	27	6	116927
5	13.4	74.4	37	6	AX453551
6	13.4	74.4	38	6	AX453553
7	13.4	74.4	47	12	SYNPRMA
8	13.4	74.4	72	11	G31952
9	13.4	74.4	81	12	SYNPRMB
10	13.4	74.4	81	12	SYNPRMC
11	13.4	74.4	81	12	SYNPRMG
12	13.4	74.4	90	12	SYNPRMH
13	13.4	74.4	94	10	F391178502
14	13.4	74.4	99	10	MUSCEREBI
15	13.2	73.3	66	9	AF043830
16	13.2	73.3	75	14	NV1243788
17	13	72.2	21	6	AX154058
18	13	72.2	22	6	A86933
19	13	72.2	22	6	AR048345
20	13	72.2	22	6	AR079236
21	13	72.2	23	6	AR118353
22	13	72.2	36	6	AR084537
23	13	72.2	43	6	AR011922
24	13	72.2	43	6	AR017793
25	13	72.2	43	6	AR077198
26	13	72.2	43	6	E38119
27	13	72.2	51	6	AX204206
28	13	72.2	89	4	MYB21F
29	12.8	71.1	18	6	A20774
30	12.8	71.1	18	6	AR095805
31	12.8	71.1	18	6	AR107992
32	12.8	71.1	18	6	AR144253
33	12.8	71.1	18	6	AR209935
34	12.8	71.1	21	6	AX404317
35	12.8	71.1	21	6	AX404318
36	12.8	71.1	24	6	A97517
37	12.8	71.1	31	6	BD002684
38	12.8	71.1	33	6	A40968
39	12.8	71.1	33	6	A44341
40	12.8	71.1	33	6	A80274
41	12.8	71.1	33	6	AR076547
42	12.8	71.1	50	9	HSU96944
43	12.8	71.1	50	9	HSU96945
44	12.8	71.1	51	6	AX160985
45	12.8	71.1	51	6	AX162476

ALIGNMENTS

RESULT 1

AF218944/c 91 bp RNA linear VRL 18-JUL-2000

LOCUS

DEFINITION

partial cds.

AF218944

AF218944.1 GI:9255784

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Normal virus.

Viruses; ssRNA positive-strand viruses, no DNA stage;

Caliciviridae; Norovirus.

REFERENCE

1 (bases 1 to 91)

AUTHORS

Nakata,S.

TITLE

Direct Submission

ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
JOURNAL	1 (bases 1 to 78)
TITLE	van Es,J.H.
FEATURES	Direct Submission submitted (12-NOV-1998) van Es J H., Immunology, University Hospital Utrecht, Heidelberglaan 101 Utrecht, 3508 GA, THE NETHERLANDS
REFERENCE	2 (bases 1 to 78)
AUTHORS	van Es,J.H., Kirkpatrick,C., van de Wetering,M., Molenaar,M., Males,A., Kuipers,J., Destree,O., Pelfer,M. and Clevers,H.
JOURNAL	Identification of APC2, a homologue of the adenomatous polyposis coli tumour suppressor
MEDLINE	Curr. Biol. 9 (2), 105-108 (1999)
PUBMED	99147086
FEATURES	10021369
SOURCE	Location/Qualifiers
gene	1..78 /organism="Mus musculus"
exon	/db_xref="taxon:10090" 1..78 /clone_lib="Mouse P1" /gene="APC2" 1..78 /product="APC2 protein" /number=11 /usedin=AJ130783.apc2_CDS
BASE COUNT	15 a 23 c 26 g 14 t
ORIGIN	
Query Match	76.7%; Score 13.8; DB 10; Length 78;
Best Local Similarity	88.2%; Pred. No. 7.3e+03;
Matches	15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY	1 CTCAGAGAGAGGAGGC 17
Db	54 CTCGACGACGCGAGGC 70
RESULT 4	
LOCUS	116927
DEFINITION	Sequence 2 from patent US 5482836.
ACCESSION	116927
VERSION	116927.1 GI:1251835
KEYWORDS	
SOURCE	Unknown.
ORGANISM	Unclassified.
REFERENCE	1 (bases 1 to 27)
AUTHORS	Cantor,C.R., Ito,T. and Smith,C.L.
TITLE	DNA purification by triplex-affinity capture and affinity capture electrophoresis
JOURNAL	Patent: US 5482836-A 2 09-JAN-1996;
FEATURES	Location/Qualifiers
SOURCE	1..27 /organism="unknown"
BASE COUNT	1 a 14 c 2 g 10 t
ORIGIN	
Query Match	74.4%; Score 13.4; DB 6; Length 27;
Best Local Similarity	93.3%; Pred. No. 1.3e+04;
Matches	14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY	4 AGAGAGAGAGGCC 18
Db	18 AGACAGACGATCC 4
RESULT 5	
RX453551	

LOCUS AX453551 37 bp DNA linear PAT 06-JUL-2002
DEFINITION Sequence 36 from Patent WO0244215.
ACCESSION AX453551
VERSION AX453551.1 GI:21712801
KEYWORDS
SOURCE . synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1
AUTHORS Rekke,O.H., Sandlie,I. and Lauvrak,V.
TITLE Product
JOURNAL Patent: WO 0244215-A 36 06-JUN-2002;
REKKE OLE HENRIK ANDRE (NO); SANDLIE INGER (NO); LAUVRAK VIGDIS (NO)
FEATURES
source location/Qualifiers
1.37 /organism="synthetic construct"
/db_xref="taxon:32630"
/note="PRIMER"
BASE COUNT 13 a 8 c 14 g 2 t
ORIGIN
Query Match 74.4%; Score 13.4; DB 6; Length 37;
Best Local Similarity 93.3%; Pred. No. 1.3e+04;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 4 AGAGAGAGAGAGGCC 18
|||||
Db 2 AGAGAGAGAGAGATCC 16
RESULT 6
AX453553 38 bp DNA linear PAT 06-JUL-2002
LOCUS Sequence 38 from Patent WO0244215.
DEFINITION AX453553
ACCESSION AX453553
VERSION AX453553.1 GI:21712803
KEYWORDS . synthetic construct.
SOURCE synthetic construct.
ORGANISM artificial sequences.
REFERENCE 1
AUTHORS Rekke,O.H., Sandlie,I. and Lauvrak,V.
TITLE Product
JOURNAL Patent: WO 0244215-A 38 06-JUN-2002;
REKKE OLE HENRIK ANDRE (NO); SANDLIE INGER (NO); LAUVRAK VIGDIS (NO)
FEATURES
source location/Qualifiers
1.38 /organism="synthetic construct"
/db_xref="taxon:32630"
/note="PRIMER"
BASE COUNT 15 a 9 c 11 g 3 t
ORIGIN
Query Match 74.4%; Score 13.4; DB 6; Length 38;
Best Local Similarity 93.3%; Pred. No. 1.3e+04;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 4 AGAGAGAGAGAGGCC 18
|||||
Db 2 AGAGAGAGAGAGATCC 16
RESULT 7
SYNPRMA 47 bp DNA linear SYN 27-APR-1993
LOCUS Artificial gene pMW701 synthetic fragment DNA.
DEFINITION M94403
ACCESSION M94403
VERSION M94403.1 GI:209149
KEYWORDS . Artificial gene DNA.
SOURCE synthetic construct
ORGANISM synthetic construct

REFERENCE 1
AUTHORS Parlewski,P., Kwinkowski,M., Wilk,A. and Klysik,J.
TITLE Dam methyltransferase sites located within the loop region of the oligopurine-oligopyrimidine sequences capable of forming H-DNA are undermethylated in vivo
JOURNAL Nucleic Acids Res. 18 (3), 605-611 (1990)
MEDLINE 90174984
PUBMED 2155405
REFERENCE 2 (bases 1 to 47)
AUTHORS Klysik,J.
TITLE Cruciform extrusion facilitates intramolecular triplex formation between distal oligopurine.oligopyrimidine tracts: long range effects
JOURNAL J. Biol. Chem. 267 (24), 17430-17437 (1992)
MEDLINE 92381066
PUBMED 1512273
FEATURES
source location/Qualifiers
1.47 /organism="synthetic construct"
/db_xref="taxon:32630"
BASE COUNT 18 a 6 c 19 g 4 t
ORIGIN
Query Match 74.4%; Score 13.4; DB 12; Length 47;
Best Local Similarity 93.3%; Pred. No. 1.3e+04;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 4 AGAGAGAGAGAGGCC 18
|||||
Db 33 AGAGAGAGAGAGATCC 47
RESULT 8
G31952/c 72 bp DNA linear STS 27-OCT-1998
LOCUS MCW92 Chicken RPMacroojmans Gallus gallus STS genomic, sequence tagged site.
DEFINITION G31952
ACCESSION G31952
VERSION G31952.1 GI:2137898
KEYWORDS STS.
SOURCE Gallus gallus.
ORGANISM Gallus gallus.
REFERENCE 1
AUTHORS Crooijmans,R.P., Dijkhof,R.J., van der Poel,J.J. and Groenen,M.A.
TITLE New microsatellite markers in chicken optimized for automated fluorescent genotyping
JOURNAL Anim. Genet. 28 (6), 427-437 (1997)
MEDLINE 98251489
PUBMED 9589584
COMMENT
Contact: Richard Crooijmans
Department of Animal Breeding
Wageningen Agricultural University
Box 338, 6700 AH Wageningen, The Netherlands
Email: richard.crooijmans@ag.vf.wau.nl
Primer A: GATCCCTCATGAAATACAGTT
Primer B: GATCGCAAAACCGCTTTGTC
STS size: 72
PCR Profile:
Presoak: 94 degrees C for 5 minute (s):
Denaturation: 94 degrees C for 30 seconds
Annealing: 55 degrees C for 45 seconds
Polymerization: 72 degrees C for 90 seconds
PCR Cycles: 35
Thermal Cycler: MJ research
Protocol:
Template: 10-100 ng
Primer: each 295 nM
dNTPs: each 200 uM

Buffer:

MgCl₂: 1.5 mM
KCl: 50 mM
Tris-HCl: 10 mM
Tetramethylammoniumchloride(TMAC): 1 mM
pH: 8.3

size range: 71-75.

FEATURES
source location/Qualifiers
1..72
/organism="Gallus gallus"
/db_xref="taxon:9031"
/clone_lib="Chicken RPMACrooljmans"

STS
primer_bind 1..21
primer_bind complement(53..72)
BASE COUNT 18 a 26 c 9 g 19 t

ORIGIN

Query Match 74.4%; Score 13.4; DB 11; Length 72;
Best Local Similarity 93.3%; Pred. No. 1.2e+04;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 4 AGAGAGAGAGAGCC 18
|||||
DB 32 AGAGAGAGAGAGAAC 18

RESULT 9
SYNPRMB 81 bp DNA linear SYN 27-APR-1993
LOCUS SYNPRMB
DEFINITION Artificial gene PRW2704 synthetic fragment DNA.
ACCESSION M94404
VERSION M94404.1 GI:209150
KEYWORDS
SOURCE Artificial gene DNA.
ORGANISM synthetic construct
artificial sequences.
REFERENCE 1 (bases 1 to 81)
AUTHORS Klysik,J.
TITLE Cruciform extrusion facilitates intramolecular triplex formation
between distal oligopurine-oligopyrimidine tracts: long range effects
JOURNAL J. Biol. Chem. 267 (24), 17430-17437 (1992)
MEDLINE 92381066
PUBMED 1512273
FEATURES
source location/Qualifiers
1..81
/organism="synthetic construct"
/db_xref="taxon:32630"

BASE COUNT 33 a 8 c 21 g 19 t

ORIGIN

Query Match 74.4%; Score 13.4; DB 12; Length 81;
Best Local Similarity 93.3%; Pred. No. 1.2e+04;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 4 AGAGAGAGAGAGCC 18
|||||
DB 67 AGAGAGAGAGAGATCC 81

RESULT 10
SYNPRMC 81 bp DNA linear SYN 27-APR-1993
LOCUS SYNPRMC
DEFINITION Artificial gene PRW2706 synthetic fragment DNA.
ACCESSION M94405
VERSION M94405.1 GI:209151
KEYWORDS
SOURCE Artificial gene DNA.

ORGANISM synthetic construct
artificial sequences.
REFERENCE 1 (bases 1 to 81)
AUTHORS Klysik,J.
TITLE Cruciform extrusion facilitates intramolecular triplex formation
between distal oligopurine-oligopyrimidine tracts: long range effects
JOURNAL J. Biol. Chem. 267 (24), 17430-17437 (1992)
MEDLINE 92381066
PUBMED 1512273
FEATURES
source location/Qualifiers
1..81
/organism="synthetic construct"
/db_xref="taxon:32630"

BASE COUNT 25 a 14 c 31 g 11 t

ORIGIN

Query Match 74.4%; Score 13.4; DB 12; Length 81;
Best Local Similarity 93.3%; Pred. No. 1.2e+04;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 4 AGAGAGAGAGAGCC 18
|||||
DB 67 AGAGAGAGAGAGATCC 81

RESULT 11
SYNPRMG 81 bp DNA linear SYN 27-APR-1993
LOCUS SYNPRMG
DEFINITION Artificial gene PRW2714 synthetic fragment DNA.
ACCESSION M94409
VERSION M94409.1 GI:209155
KEYWORDS
SOURCE Artificial gene DNA.
ORGANISM synthetic construct
artificial sequences.
REFERENCE 1 (bases 1 to 81)
AUTHORS Klysik,J.
TITLE Cruciform extrusion facilitates intramolecular triplex formation
between distal oligopurine-oligopyrimidine tracts: long range effects
JOURNAL J. Biol. Chem. 267 (24), 17430-17437 (1992)
MEDLINE 92381066
PUBMED 1512273
FEATURES
source location/Qualifiers
1..81
/organism="synthetic construct"
/db_xref="taxon:32630"

BASE COUNT 27 a 14 c 27 g 13 t

ORIGIN

Query Match 74.4%; Score 13.4; DB 12; Length 81;
Best Local Similarity 93.3%; Pred. No. 1.2e+04;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 4 AGAGAGAGAGAGCC 18
|||||
DB 67 AGAGAGAGAGAGATCC 81

RESULT 12
SYNPRWH 90 bp DNA linear SYN 27-APR-1993
LOCUS SYNPRWH
DEFINITION Artificial gene PRW2716 synthetic fragment DNA.
ACCESSION M94410
VERSION M94410.1 GI:209156
KEYWORDS
SOURCE Artificial gene DNA.
ORGANISM synthetic construct
artificial sequences.
REFERENCE 1 (bases 1 to 90)
AUTHORS Klysik,J.
TITLE Cruciform extrusion facilitates intramolecular triplex formation

between distal oligopurine-oligopyrimidine tracts: long range effects

JOURNAL J. Biol. Chem. 267 (24), 17430-17437 (1992)
MEDLINE 92381066
PUBMED 1512273

FEATURES
source location/Qualifiers
1..90
/organism="synthetic construct"
/db_xref="taxon:32630"

BASE COUNT 26 a 13 c 27 g 24 t

ORIGIN

Query Match 74.4%; Score 13.4; DB 12; Length 90;
Best Local Similarity 93.3%; Pred. No. 1.2e+04;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 4 AGAGAGAGAGAGGCC 18
|||||
Db 76 AGAGAGAGAGAGATCC 90

RESULT 13
F391178S02 94 bp DNA linear ROD 13-DEC-2001

LOCUS F391178S02
DEFINITION Mus musculus F-box Wp40 repeat protein 1 (Fbxw1) gene, exon 2.
ACCESSION AF391179
VERSION AF391179.1 GI:17646199

KEYWORDS
SEGMENT 2 of 14
SOURCE Mus musculus.
ORGANISM Mus musculus.
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 94)
AUTHORS Maruyama Si.S., Hatakeyama,S., Nakayama,K., Ishida,N., Kawakami,K. and Nakayama,K.I.
TITLE Characterization of a Mouse Gene (Fbxw6) That Encodes a Homologue of Catenohaditils elegans SEU-10
PUBMED 11735228
REFERENCE 2 (bases 1 to 94)
AUTHORS Hatakeyama,S.
TITLE Direct Submission
JOURNAL Submitted (13-JUN-2001) Department of Molecular and Cellular Biology, Medical Institute of Bioregulation, Kyushu University, 3-1-1, Maidashi, Higashi-ku, Fukuoka-shi 812-8582, Japan

FEATURES
source location/Qualifiers
1..94
/organism="Mus musculus"
/strain="129/Sv"
/db_xref="taxon:10090"
9..86
/gene="Fbxw1"
/number=2

BASE COUNT 37 a 18 c 20 g 19 t

ORIGIN

Query Match 74.4%; Score 13.4; DB 10; Length 94;
Best Local Similarity 93.3%; Pred. No. 1.2e+04;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 CTCGAGAGAGAGA 15
|||||
Db 14 CTCAGAGAGAGAGA 28

RESULT 14
MUSCEREBI 99 bp mRNA linear ROD 31-JUL-1995

LOCUS MUSCEREBI
DEFINITION Mus musculus (clone Gcap15) mRNA sequence.
ACCESSION L10902
VERSION L10902.1 GI:192551

KEYWORDS

SOURCE Mus musculus (library: lambda Zap) neonate cerebellum cDNA to mRNA.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 99)
REFERENCE Identification of novel cDNAs expressed in murine cerebellum
AUTHORS Kamouris,M.
TITLE Thesis (1993)
JOURNAL 2 (bases 1 to 99)
REFERENCE Kamouris,M., Ph.D., Sangameswaran,L., Tzitarhou,L.C., Luo,F., Dlouhy,S.R., Ghetti,B. and Hodges,M.E.
AUTHORS Novel cDNA clones obtained by antibody screening of a mouse cerebellar cDNA expression library
JOURNAL Unpublished (1993)
FEATURES location/Qualifiers
1..99
/organism="Mus musculus"
/db_xref="taxon:10090"
/haploype="PI wv/+"
/tissue_type="cerebellum"
/dev_stage="neonate"
/tissue_lib="lambda Zap"

BASE COUNT 22 a 16 c 47 g 14 t

ORIGIN

Query Match 74.4%; Score 13.4; DB 10; Length 99;
Best Local Similarity 93.3%; Pred. No. 1.2e+04;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 TCAGAGAGAGAGAC 16
|||||
Db 4 TCCGAGAGAGAGAG 18

RESULT 15
AF043830 66 bp mRNA linear PRI 11-NOV-1998

LOCUS AF043830
DEFINITION Homo sapiens patient CS-2 clone 34 T cell receptor beta chain CDR3 (TCRB) mRNA, partial cds.
ACCESSION AF043830
VERSION AF043830.1 GI:3859289

KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo. 1 (bases 1 to 66)
REFERENCE Striebig,C.C., Falta,M.T., Wang,Y., Bill,J. and Kotzin,B.L.
TITLE Selective accumulation of related CD4+ T cell clones in the synovial fluid of patients with Rheumatoid arthritis
JOURNAL J. Immunol. 161 (8), 4428-4436 (1998)
MEDLINE 98451502
PUBMED 9780222

REFERENCE 2 (bases 1 to 66)
AUTHORS Striebig,C.C., Falta,M.T., Wang,Y., Bill,J. and Kotzin,B.L.
TITLE Direct Submission
JOURNAL Submitted (21-JAN-1998) Pediatrics, National Jewish Medical and Research Center, 1400 Jackson St., Denver, CO 80206, USA

FEATURES
source location/Qualifiers
1..66
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="34"
/tissue_type="peripheral blood"
/note="from rheumatoid arthritis patient CS-2 (TCRB17) [TCRB155]"
[TCRB155]
<1..>66
/gene="TCRB"
<1..>66
/gene="TCRB"
/note="Contains the 3' end of TCRB, the ndn, and the 5' end of the TCRB."
/codon_start=1

/product="T cell receptor beta chain CDR3"
/protein_id="AAC72630.1"
/db_xref="GI:3859290"
/translation="AOKNPTAFYLCASQSDPQHFG"
BASE COUNT 17 a 22 c 13 g 14 t
ORIGIN

Query Match 73.3%; Score 13.2; DB 9; Length 66;
Best Local Similarity 83.3%; Pred. No. 1.6e+04;
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 CTCAGAGAGAGAGAGCC 18
|||||
Db 36 CTCACAGAGAGACGCC 53

Search completed: November 9, 2002, 14:59:29
Job time : 2821 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 9, 2002, 12:28:01 ; Search time 296 Seconds

(without alignments)
136.946 Million cell updates/sec

Title: US-09-902-772-1_COPY_645_662

Perfect score: 18

Sequence: 1 cctagagagagagagagc 18

Scoring table:

IDENTITY_NTC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 2390332

Minimum DB seq length: 0

Maximum DB seq length: 100

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```
1: N_Genes2_101002:*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:*
25: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	15.4	85.6	65	24	ABN30396 Rat spliced trans
2	14	77.8	60	24	ABN40086 Human spliced tran
3	14	77.8	60	24	ABN45059 Human spliced tran
4	13.8	76.7	17	24	ABN02627 Human GDMLP-1 17-m
5	13.8	76.7	25	24	ABN05556 Human GDMLP-1 25-m
6	13.8	76.7	25	24	ABN05557 Human GDMLP-1 25-m
7	13.8	76.7	25	24	ABN05558 Human GDMLP-1 25-m
8	13.8	76.7	25	24	ABN05559 Human GDMLP-1 25-m
9	13.8	76.7	25	24	ABN05560 Human GDMLP-1 25-m

10	13.8	76.7	25	24	ABN05561 Human GDMLP-1 25-m
11	13.8	76.7	25	24	ABN05562 Human GDMLP-1 25-m
12	13.8	76.7	25	24	ABN05563 Human GDMLP-1 25-m
13	13.8	76.7	25	24	ABN05564 Human GDMLP-1 25-m
14	13.4	74.4	20	24	AA039624 Human SR-cyp antis
15	13.4	74.4	21	22	AA062255 SLC16 polymorphis
16	13.4	74.4	27	17	AA03688 Triplex-affinity D
17	13.4	74.4	37	24	AB075846 PCR primer FCRIFOR
18	13.4	74.4	37	24	AB075848 PCR primer FCR-Bam
19	13.4	74.4	44	15	AA078847 HCMV Ab light chain
20	13.4	74.4	60	24	ABN04040 Human spliced tran
21	13.4	74.4	61	21	AA015530 Human secreted pro
22	13.4	74.4	63	21	AA015545 Human secreted pro
23	13.4	74.4	65	24	ABN58012 Mouse spliced tran
24	13.4	74.4	99	22	ABA48277 Human breast cell
25	13.4	74.4	99	22	ABA66159 Human foetal liver
26	13.4	74.4	99	22	ABA33228 Probe #11694 for g
27	13.4	74.4	99	22	AAK14578 Human brain expres
28	13.4	74.4	99	22	AAK10324 Human bone marrow
29	13.4	74.4	99	22	AAI21082 Human genome-deriv
30	13.4	74.4	99	22	AAI66802 NFI related PCR P
31	13.4	74.4	99	22	AAI66343 Human clone c94397
32	13.4	74.4	99	24	AB014344 Mouse spliced tran
33	13.2	73.3	31	22	AAI68326 Downstream sequenc
34	13.2	73.3	51	21	AAAT7247 Human SCYR24 ASO p
35	13.2	73.3	65	13	AA033868 Human biallelic ma
36	13.2	73.3	69	13	AA033868 CDNA synthesis pri
37	13.2	73.3	80	13	AA034153 PCR primer EGRII-6
38	13.2	73.3	15	24	AB057606 Oligonucleotide id
39	13	72.2	21	21	AAZ76005 primer used in a n
40	13	72.2	22	18	AAAT5373 Human DNA containi
41	13	72.2	22	19	AAV59955 Human GDMLP-1 17-m
42	13	72.2	23	14	AA039034
43	13	72.2	43	20	AA081695
44	13	72.2	51	22	AAH76987
45	12.8	71.1	17	24	ABN02626

ALIGNMENTS

```
RESULT 1
ABN30396
ID ABN30396 standard; DNA; 65 BP.
XX
AC ABN30396;
XX
DT 15-JUL-2002 (first entry)
XX
DE Rat spliced transcript detection oligonucleotide seq ID NO:3144.
XX
KW Human; mouse; rat; splice transcript; detection; RNA transcript;
XX splice variant; transcriptome; oligonucleotide library; ss.
XX
OS Rattus norvegicus.
XX
PN WO200210449-A2.
XX
PD 07-FEB-2002.
XX
PE 20-JUL-2001; 2001WO-IB01903.
XX
PR 28-JUL-2000; 2000US-221607P.
XX
PR 02-MAY-2001; 2001US-287724P.
XX
PA (COMP-) COMUGEN INC.
XX
PI Shoshan A, Wasserman A, Mintz E, Mintz L, Faigler S;
XX WPL; 2002-257383/30.
XX
PT New oligonucleotide libraries comprising oligonucleotides which
selectively hybridize to mRNAs transcribed from a transcription unit of
```

```
PT a genome, useful for detecting tissue-, pathology-, and
PT developmental-specific genes
XX
PS Example 1; SEQ ID 3144; 47pp; English.
XX
CC The present invention describes oligonucleotide libraries for detecting
CC messenger RNAs that populate a (sub-)transcriptome, where the
CC (sub-)transcriptome comprises messenger RNAs transcribed from multiple
CC transcription units that populate a genome. The library comprises
CC several oligonucleotides, each capable of hybridizing selectively to a
CC set of messenger RNAs transcribed from a given transcription unit of
CC the genome, which encodes one or more messenger RNA splice variants.
CC The oligonucleotide libraries are useful for detecting mRNAs from a
CC biological sample, in expression profiling studies, in qualitatively or
CC quantitatively characterizing the corresponding transcriptome, and in
CC detecting RNA transcripts and splice variants of human or animal
CC transcriptomes. The libraries may also be used as specialised mini
CC libraries to detect transcripts of a sub-transcriptome under a
CC particular biological or pathological state, and so allowing the
CC detection of tissue- and pathology-specific genes such as those genes
CC only expressed in specific tissue under a specific pathological
CC condition; to detect developmental specific genes; and to detect RNA
CC transcripts and splice variants of a transcriptome of a patient suffering
CC from a particular disorder. ABN27253 to ABN59589 represent
CC oligonucleotide sequences from rats, humans and mice, which are used in
CC the exemplification of the present invention.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 65 BP; 21 A; 16 C; 16 G; 12 T; 0 other;
Query Match 85.6%; Score 15.4; DB 24; Length 65;
Best Local Similarity 94.1%; Pred. No. 6e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 TCAGAGAGAGAGAGCC 18
DB 19 TCAGAGAGAGAGAGCC 35
RESULT 2
ABN40086/c
ID ABN40086 standard; DNA; 60 BP.
XX
AC ABN40086;
XX
DT 15-JUL-2002 (first entry)
XX
DE Human spliced transcript detection oligonucleotide SEQ ID NO:12834.
XX
KW Human; mouse; rat; splice transcript; detection; RNA transcript;
KW splice variant; transcriptome; oligonucleotide library; ss.
XX
OS Homo sapiens.
XX
PN WO200210449-A2.
XX
PD 07-FEB-2002.
XX
PF 20-JUL-2001; 2001WO-1B01903.
XX
PR 28-JUL-2000; 2000US-221607P.
PR 02-MAY-2001; 2001US-287724P.
XX
PA (COMP-) COMPUGEN INC.
XX
PI Shoshan A, Wasserman A, Mintz E, Mintz L, Faigler S;
XX
DR WPI; 2002-257383/30.
XX
PT New oligonucleotide libraries comprising oligonucleotides which
PT selectively hybridize to mRNAs transcribed from a transcription unit of
```

```
PT a genome, useful for detecting tissue-, pathology-, and
PT developmental-specific genes
XX
PS Example 1; SEQ ID 12834; 47pp; English.
XX
CC The present invention describes oligonucleotide libraries for detecting
CC messenger RNAs that populate a (sub-)transcriptome, where the
CC (sub-)transcriptome comprises messenger RNAs transcribed from multiple
CC transcription units that populate a genome. The library comprises
CC several oligonucleotides, each capable of hybridizing selectively to a
CC set of messenger RNAs transcribed from a given transcription unit of
CC the genome, which encodes one or more messenger RNA splice variants.
CC The oligonucleotide libraries are useful for detecting mRNAs from a
CC biological sample, in expression profiling studies, in qualitatively or
CC quantitatively characterizing the corresponding transcriptome, and in
CC detecting RNA transcripts and splice variants of human or animal
CC transcriptomes. The libraries may also be used as specialised mini
CC libraries to detect transcripts of a sub-transcriptome under a
CC particular biological or pathological state, and so allowing the
CC detection of tissue- and pathology-specific genes such as those genes
CC only expressed in specific tissue under a specific pathological
CC condition; to detect developmental specific genes; and to detect RNA
CC transcripts and splice variants of a transcriptome of a patient suffering
CC from a particular disorder. ABN27253 to ABN59589 represent
CC oligonucleotide sequences from rats, humans and mice, which are used in
CC the exemplification of the present invention.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 60 BP; 5 A; 25 C; 8 G; 22 T; 0 other;
Query Match 77.8%; Score 14; DB 24; Length 60;
Best Local Similarity 100.0%; Pred. No. 2.5e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 4 AGAGAGAGAGAGAC 17
DB 47 AGAGAGAGAGAGAC 34
RESULT 3
ABN45059/c
ID ABN45059 standard; DNA; 60 BP.
XX
AC ABN45059;
XX
DT 15-JUL-2002 (first entry)
XX
DE Human spliced transcript detection oligonucleotide SEQ ID NO:17807.
XX
KW Human; mouse; rat; splice transcript; detection; RNA transcript;
KW splice variant; transcriptome; oligonucleotide library; ss.
XX
OS Homo sapiens.
XX
PN WO200210449-A2.
XX
PD 07-FEB-2002.
XX
PF 20-JUL-2001; 2001WO-1B01903.
XX
PR 28-JUL-2000; 2000US-221607P.
PR 02-MAY-2001; 2001US-287724P.
XX
PA (COMP-) COMPUGEN INC.
XX
PI Shoshan A, Wasserman A, Mintz E, Mintz L, Faigler S;
XX
DR WPI; 2002-257383/30.
XX
PT New oligonucleotide libraries comprising oligonucleotides which
PT selectively hybridize to mRNAs transcribed from a transcription unit of
```

```

PT  a genome, useful for detecting tissue-, pathology-, and
PT  developmental-specific genes
PS
PS  Example 1; SEQ ID 17807; 47pp; English.
XX
CC  The present invention describes oligonucleotide libraries for detecting
CC  messenger RNAs that populate a (sub-)transcriptome, where the
CC  (sub-)transcriptome comprises messenger RNAs transcribed from multiple
CC  transcription units that populate a genome. The library comprises
CC  several oligonucleotides, each capable of hybridizing selectively to a
CC  set of messenger RNAs transcribed from a given transcription unit of
CC  the genome, which encodes one or more messenger RNA splice variants.
CC  The oligonucleotide libraries are useful for detecting mRNAs from a
CC  biological sample, in expression profiling studies, in qualitatively or
CC  quantitatively characterizing the corresponding transcriptome, and in
CC  detecting RNA transcripts and splice variants of human or animal
CC  transcriptomes. The libraries may also be used as specialised mini
CC  libraries to detect transcripts of a sub-transcriptome under a
CC  particular biological or pathological state, and so allowing the
CC  detection of tissue- and pathology-specific genes such as those genes
CC  only expressed in specific tissue under a specific pathological
CC  condition; to detect developmental specific genes; and to detect RNA
CC  transcripts and splice variants of a transcriptome of a patient suffering
CC  from a particular disorder. ABN27253 to ABN9589 represent
CC  oligonucleotide sequences from rats, humans and mice, which are used in
CC  the exemplification of the present invention.
CC  N.B. The sequence data for this patent did not form part of the printed
CC  specification, but was obtained in electronic format directly from WIPO
CC  at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ  Sequence 60 BP; 7 A; 17 C; 13 G; 23 T; 0 other;
XX
Query Match          77.8%; Score 14; DB 24; Length 60;
Best Local Similarity 100.0%; Pred. NO. 2.5e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0
QY      1 CTCAGAGAGAGAGG 14
        |||||||||||
Db       21 CTCAGAGAGAGAGG 8
XX
RESULT 4
ABN02627
ID      ABN02627 standard; DNA; 17 BP.
XX
AC      ABN02627;
XX
AD
XX
DT      29-MAY-2002 (first entry)
XX
DE      Human GDMF-1 17-mer scanning SEQ ID NO:4 sequence SEQ ID NO:2619.
XX
KW      Human; genome-derived myosin-like protein 1; GDMF-1; hGDMF-1; heart;
KW      muscle; myosin; chromosome 22; gene therapy; vaccine; heart disease;
KW      skeletal muscle disorder; amplicon; screening; ss.
XX
OS      Homo sapiens.
XX
XX      WO200192524-A2.
XX
XX      06-DEC-2001.
XX
XX      25-MAY-2001; 2001WO-US16981.
XX
XX      26-MAY-2000; 2000US-207456P.
XX      21-SEP-2000; 2000US-234687P.
XX      27-SEP-2000; 2000US-236359P.
XX      04-OCT-2000; 2000GB-0024263.
XX      30-JAN-2001; 2001WO-US00661.
XX      30-JAN-2001; 2001WO-US00662.
XX      30-JAN-2001; 2001WO-US00663.
XX      30-JAN-2001; 2001WO-US00664.
XX      30-JAN-2001; 2001WO-US00665.
XX      30-JAN-2001; 2001WO-US00666.
XX

```

PR	30-JAN-2001;	2001WO-US00667.
PR	30-JAN-2001;	2001WO-US00668.
PR	30-JAN-2001;	2001WO-US00669.
PR	30-JAN-2001;	2001WO-US00670.
PR	05-FEB-2001;	2001US-266860P.
XX	(AEOM-) AEOMICA INC.	
XX		
PI	Gu Y, Ji Y, Penn SG, Hanzel DK, Rank DR, Chen W, Shannon ME;	
XX		
DR	WPI; 2002-179446/23.	
XX		
PT	New polypeptide, for raising antibodies that recognize hGDMLP-1	
PT	proteins, or as specific biomolecule capture probes for	
PT	surface-enhanced laser desorption/ionization, comprises human	
PT	myosin-like protein hGDMLP-1 -	
PS		
PS	Disclosure; SEQ ID 2619; 214pp; English.	
XX		
CC	The present invention describes a human genome-derived myosin-like	
CC	protein 1 (hGDMLP-1). The protein and polynucleotide sequences of	
CC	hGDMLP-1 can be used in gene therapy and vaccine production. The	
CC	hGDMLP-1 nucleic acids can be used as probes to detect, characterise	
CC	and quantify hGDMLP-1 nucleic acids in samples, as amplification	
CC	substrates, to provide initial substrates for the recombinant engineering	
CC	of hGDMLP-1 protein variants having desired phenotypic improvements, and	
CC	for expressing the proteins. The hGDMLP-1 proteins or polypeptides may	
CC	be used as immunogens to raise antibodies that specifically recognise	
CC	hGDMLP-1 proteins, as standards in assays used to determine the	
CC	concentration and/or amount specifically of hGDMLP proteins, as specific	
CC	biomolecule capture probes for surface-enhanced laser desorption	
CC	ionisation, as therapeutic supplement in patients having specific	
CC	deficiency in hGDMLP-1 production, and in vaccines or for replacement	
CC	therapy. The polynucleotide sequences encoding hGDMLP-1 may be used for	
CC	diagnosing a disorder associated with the expression of hGDMLP-1, in	
CC	a particular heart and skeletal muscle disorders. hGDMLP-1 is localised to	
CC	chromosome 22. The present sequence represents an oligomer used in the	
CC	screening of the hGDMLP-1 sequence in the exemplification of the present	
CC	invention.	
CC	N.B. The sequence data for this patent did not form part of the printed	
CC	specification, but was obtained in electronic format directly from WIPO	
CC	at ftp.wipo.int/pub/published_pcl_sequence.	
XX		
SQ	Sequence 17 BP; 5 A; 5 C; 6 G; 1 T; 0 other:	
	Query Match 76.7%; Score 13.8; DB 24; Length 17;	
	Best Local Similarity 88.2%; Pred. No. 2.9e+03;	
	Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0,	
QY	1 CTCAGAGGAGAGGAGC 17	
DB	1 CTCAGAGGAGCGAGCAGC 17	
	RESULT 5	
ID	ABN05556	
XX	ABN05556 standard; DNA; 25 BP.	
XX		
AC	ABN05556;	
DT	29-MAY-2002 (first entry)	
DE		
XX	Human GDMLP-1 25-mer scanning SEQ ID NO:4 sequence SEQ ID NO:5548.	
KM	Human; genome-derived myosin-like protein 1; GDMLP-1; hGDMLP-1; heart;	
KW	muscle; myosin; chromosome 22; gene therapy; vaccine; heart disease;	
XX	skeletal muscle disorder; amplicon; screening; ss.	
OS	Homo sapiens.	
XX		
NN	WO200192524-A2.	
D		
06-DEC-2001		

XX XX 25-MAY-2001; 2001WO-US16981.
XX
PR 26-MAY-2000; 2000US-207456P.
PR 21-SEP-2000; 2000US-234687P.
PR 27-SEP-2000; 2000US-236359P.
PR 04-OCT-2000; 2000GB-0024263.
PR 30-JAN-2001; 2001MO-US00661.
PR 30-JAN-2001; 2001MO-US00662.
PR 30-JAN-2001; 2001MO-US00663.
PR 30-JAN-2001; 2001MO-US00664.
PR 30-JAN-2001; 2001MO-US00665.
PR 30-JAN-2001; 2001MO-US00666.
PR 30-JAN-2001; 2001MO-US00667.
PR 30-JAN-2001; 2001MO-US00668.
PR 30-JAN-2001; 2001MO-US00669.
PR 30-JAN-2001; 2001MO-US00670.
PR 05-FEB-2001; 2001US-266860P.
XX
PA (AEOM-) AEOMICA INC.
XX
PI Gu Y, Ji Y, Penn SG, Hanzel DK, Rank DR, Chen W, Shannon ME;
XX
DR WPI: 2002-179446/23.
XX
XX New polypeptide, for raising antibodies that recognize hGDMRP-1
PT proteins, or as specific biomolecule capture probes for
PT surface-enhanced laser desorption/ionization, comprises human
PX myosin-like protein hGDMRP-1 -
PT
PS Disclosure: SEQ ID 5548; 214pp; English.
XX
XX The present invention describes a human genome-derived myosin-like
CC protein 1 (hGDMRP-1). The protein and polynucleotide sequences of
CC hGDMRP-1 can be used in gene therapy and vaccine production. The
CC hGDMRP-1 nucleic acids can be used as probes to detect, characterise
CC and quantify hGDMRP-1 nucleic acids in samples, as amplification
CC substrates, to provide initial substrates for the recombinant engineering
CC of hGDMRP-1 protein variants having desired phenotypic improvements, and
CC for expressing the proteins. The hGDMRP-1 proteins or polypeptides may
CC be used as immunogens to raise antibodies that specifically recognise
CC hGDMRP-1 proteins, as standards in assays used to determine the
CC concentration and/or amount specifically of hGDMRP proteins, as specific
CC biomolecule capture probes for surface-enhanced laser desorption/
CC ionisation, as therapeutic supplement in patients having specific
CC deficiency in hGDMRP-1 production, and in vaccines or for replacement
CC therapy. The polynucleotide sequences encoding hGDMRP-1 may be used for
CC diagnosing a disorder associated with the expression of hGDMRP-1, in
CC particular heart and skeletal muscle disorders. hGDMRP-1 is localised to
CC chromosome 22. The present sequence represents an oligomer used in the
CC screening of the hGDMRP-1 sequence in the exemplification of the present
CC invention.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequence.
XX

Sequence 25 BP; 6 A; 7 C; 9 G; 3 T; 0 other;

Query Match 76.7%; Score 13.8; DB 24; Length 25;
Best Local Similarity 88.2%; Pred. No. 3e+03;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0

QY 1 CTCACAGACGAGCAGC 17
||| ||||||| ||| |||
Db 9 CTCACAGACGAGCAGCAGC 25

RESULT 6
ABN05557
ID ABN05557 standard; DNA; 25 BP.

AC ABN05557;
XX

DE	29-MAY-2002	(first entry)	
XX			
DE	Human GDMLP-1 25-mer scanning SEQ ID NO:4	sequence SEQ ID NO:5549.	
XX			
KW	Human; genome-derived myosin-like protein 1; GDMLP-1; hGDMLP-1; heart;		
KM	muscle; myosin; chromosome 22; gene therapy; vaccine; heart disease;		
KW	skeletal muscle disorder; amplicon; screening; ss.		
XX			
OS	Homo sapiens.		
XX			
PN	WO200192524-A2.		
XX			
PD	06-DEC-2001.		
XX			
PF	25-MAY-2001; 2001WO-US16981.		
XX			
PR	26-MAY-2000; 2000US-207456P.		
PR	21-SEP-2000; 2000US-234687P.		
PR	27-SEP-2000; 2000US-236359P.		
PR	04-OCT-2000; 2000GB-0024263.		
PR	30-JAN-2001; 2001WO-US00661.		
PR	30-JAN-2001; 2001WO-US00662.		
PR	30-JAN-2001; 2001WO-US00663.		
PR	30-JAN-2001; 2001WO-US00664.		
PR	30-JAN-2001; 2001WO-US00665.		
PR	30-JAN-2001; 2001WO-US00666.		
PR	30-JAN-2001; 2001WO-US00667.		
PR	30-JAN-2001; 2001WO-US00668.		
PR	30-JAN-2001; 2001WO-US00669.		
PR	30-JAN-2001; 2001WO-US00670.		
PR	05-FEB-2001; 2001US-268650P.		
XX			
PA	(AEOM-) AEOMICA INC.		
XX			
PI	Gu Y, Ji Y, Penn SG, Hanzel DK, Rank DR, Chen W, Shannon ME;		
XX			
DR	WPI: 2002-179446/23.		
XX			
PT	New polypeptide, for raising antibodies that recognize hGDMLP-1		
PT	proteins, or as specific biomolecule capture probes for		
PT	surface-enhanced laser desorption/ionization, comprises human		
PT	myosin-like protein hGDMLP-1 -		
XX			
PS	Disclosure; SEQ ID 5549; 214pp; English.		
XX			
CC	The present invention describes a human genome-derived myosin-like		
CC	protein 1 (hGDMLP-1). The protein and polynucleotide sequences of		
CC	hGDMLP-1 can be used in gene therapy and vaccine production. The		
CC	hGDMLP-1 nucleic acids can be used as probes to detect, characterise		
CC	and quantify hGDMLP-1 nucleic acids in samples, as amplification		
CC	substrates, to provide initial substrates for the recombinant engineering		
CC	of hGDMLP-1 protein variants having desired phenotypic improvements, and		
CC	for expressing the proteins. The hGDMLP-1 proteins or polypeptides may		
CC	be used as immunogens to raise antibodies that specifically recognise		
CC	hGDMLP-1 proteins, as standards in assays used to determine the		
CC	concentration and/or amount specifically of hGDMLP proteins, as specific		
CC	biomolecule capture probes for surface-enhanced laser desorption		
CC	ionisation, as therapeutic supplement in patients having specific		
CC	deficiency in hGDMLP-1 production, and in vaccines or for replacement		
CC	therapy. The polynucleotide sequences encoding hGDMLP-1 may be used for		
CC	diagnosing a disorder associated with the expression of hGDMLP-1, in		
CC	particular heart and skeletal muscle disorders. hGDMLP-1 is localised to		
CC	chromosome 22. The present sequence represents an oligomer used in the		
CC	screening of the hGDMLP-1 sequence in the exemplification of the present		
CC	invention.		
CC	N.B. The sequence data for this patent did not form part of the printed		
CC	specification, but was obtained in electronic format directly from WIPO		
CC	at ftp.wipo.int/pub/published_pot_sequence.		
XX			
SO	Sequence 25 BP; 6 A; 6 C; 10 G; 3 T; 0 other;		
XX			
Query Match	76.7%;	Score 13.8;	DB 24; Length 25;
Best Local Similarity	88.2%;	Pred No. 36+03;	

Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTCAGAGAGAGAGAGC 17
||||||| ||| |||
Db 8 CTCAGAGAGCGAGCAGC 24

RESULT 7
ABN05558
ID ABN05558 standard; DNA: 25 BP.
XX
AC ABN05558;
XX
DT 29-MAY-2002 (first entry)
XX
DE Human GDMLP-1 25-mer scanning SEQ ID NO:4 sequence SEQ ID NO:5550.
XX
KW Human; genome-derived myosin-like protein 1; GDMLP-1; hGDMLP-1; heart;
KW muscle; myosin; chromosome 22; gene therapy; vaccine; heart disease;
KW skeletal muscle disorder; amplicon; screening; ss.
XX
OS Homo sapiens.
XX
PN WO200192524-A2.
XX
PD 06-DEC-2001.
XX
PF 25-MAY-2001; 2001WO-US16981.
XX
PR 26-MAY-2000; 2000US-207456P.
PR 21-SEP-2000; 2000US-234687P.
PR 27-SEP-2000; 2000US-236359P.
PR 04-OCT-2000; 2000GB-0024263.
PR 30-JAN-2001; 2001WO-US00661.
PR 30-JAN-2001; 2001WO-US00662.
PR 30-JAN-2001; 2001WO-US00663.
PR 30-JAN-2001; 2001WO-US00664.
PR 30-JAN-2001; 2001WO-US00665.
PR 30-JAN-2001; 2001WO-US00666.
PR 30-JAN-2001; 2001WO-US00667.
PR 30-JAN-2001; 2001WO-US00668.
PR 30-JAN-2001; 2001WO-US00669.
PR 30-JAN-2001; 2001WO-US00670.
PR 05-FEB-2001; 2001US-266860P.
XX
XX (AEOM-) AEOMICA INC.
XX
PI Gu Y, Ji Y, Penn SG, Hanzel DK, Rank DR, Chen W, Shannon ME;
XX
DR WPI; 2002-179446/23.
XX
PT New polypeptide, for raising antibodies that recognize hGDMLP-1
PT proteins, or as specific biomolecule capture probes for
PT surface-enhanced laser desorption/ionization, comprises human
PT myosin-like protein hGDMLP-1 -
XX
PS Disclosure; SEQ ID 5550; 214pp; English.
XX
XX The present invention describes a human genome-derived myosin-like
CC protein 1 (hGDMLP-1). The protein and polynucleotide sequences of
CC hGDMLP-1 can be used in gene therapy and vaccine production. The
CC hGDMLP-1 nucleic acids can be used as probes to detect, characterise
CC and quantify hGDMLP-1 nucleic acids in samples, as amplification
CC substrates, to provide initial substrates for the recombinant engineering
CC of hGDMLP-1 protein variants having desired phenotypic improvements, and
CC for expressing the proteins. The hGDMLP-1 proteins or polypeptides may
CC be used as immunogens to raise antibodies that specifically recognise
CC hGDMLP-1 proteins, as standards in assays used to determine the
CC concentration and/or amount specifically of hGDMLP proteins, as specific
CC biomolecule capture probes for surface-enhanced laser desorption
CC ionization, as therapeutic supplement in patients having specific
CC deficiency in hGDMLP-1 production, and in vaccines or for replacement
CC therapy. The polynucleotide sequences encoding hGDMLP-1 may be used for

CC diagnosing a disorder associated with the expression of hGDMLP-1, in
CC particular heart and skeletal muscle disorders. hGDMLP-1 is localised to
CC chromosome 22. The present sequence represents an oligomer used in the
CC screening of the hGDMLP-1 sequence in the exemplification of the present
CC invention.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pft_sequence.
XX
SQ Sequence 25 BP; 6 A; 6 C; 11 G; 2 T; 0 other;
XX
Query Match 76.7%; Score 13.8; DB 24; Length 25;
Best Local Similarity 88.2%; Pred. No. 3e+03;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTCAGAGAGAGAGAGC 17
||||||| ||| |||
Db 7 CTCAGAGAGCGAGCAGC 23

RESULT 8
ABN05559
ID ABN05559 standard; DNA: 25 BP.
XX
AC ABN05559;
XX
DT 29-MAY-2002 (first entry)
XX
DE Human GDMLP-1 25-mer scanning SEQ ID NO:4 sequence SEQ ID NO:5551.
XX
KW Human; genome-derived myosin-like protein 1; GDMLP-1; hGDMLP-1; heart;
KW muscle; myosin; chromosome 22; gene therapy; vaccine; heart disease;
KW skeletal muscle disorder; amplicon; screening; ss.
XX
OS Homo sapiens.
XX
PN WO200192524-A2.
XX
PD 06-DEC-2001.
XX
PF 25-MAY-2001; 2001WO-US16981.
XX
PR 26-MAY-2000; 2000US-207456P.
PR 21-SEP-2000; 2000US-234687P.
PR 27-SEP-2000; 2000US-236359P.
PR 04-OCT-2000; 2000GB-0024263.
PR 30-JAN-2001; 2001WO-US00661.
PR 30-JAN-2001; 2001WO-US00662.
PR 30-JAN-2001; 2001WO-US00663.
PR 30-JAN-2001; 2001WO-US00664.
PR 30-JAN-2001; 2001WO-US00665.
PR 30-JAN-2001; 2001WO-US00666.
PR 30-JAN-2001; 2001WO-US00667.
PR 30-JAN-2001; 2001WO-US00668.
PR 30-JAN-2001; 2001WO-US00669.
PR 30-JAN-2001; 2001WO-US00670.
PR 05-FEB-2001; 2001US-266860P.
XX
XX (AEOM-) AEOMICA INC.
XX
PI Gu Y, Ji Y, Penn SG, Hanzel DK, Rank DR, Chen W, Shannon ME;
XX
DR WPI; 2002-179446/23.
XX
PT New polypeptide, for raising antibodies that recognize hGDMLP-1
PT proteins, or as specific biomolecule capture probes for
PT surface-enhanced laser desorption/ionization, comprises human
PT myosin-like protein hGDMLP-1 -
XX
PS Disclosure; SEQ ID 5551; 214pp; English.
XX
XX The present invention describes a human genome-derived myosin-like
CC protein 1 (hGDMLP-1). The protein and polynucleotide sequences of

CC hGDMRP-1 nucleic acids can be used in gene therapy and vaccine production. The
 CC hGDMRP-1 nucleic acids can be used as probes to detect, characterise
 CC and quantify hGDMRP-1 nucleic acids in samples, as amplification
 CC substrates, to provide initial substrates for the recombinant engineering
 CC of hGDMRP-1 protein variants having desired phenotypic improvements, and
 CC for expressing the proteins. The hGDMRP-1 proteins or polypeptides may
 CC be used as immunogens to raise antibodies that specifically recognise
 CC hGDMRP-1 proteins, as standards in assays used to determine the
 CC concentration and/or amount specifically of hGDMRP proteins, as specific
 CC biomolecule capture probes for surface-enhanced laser desorption
 CC ionisation, as therapeutic supplement in patients having specific
 CC deficiency in hGDMRP-1 production, and in vaccines or for replacement
 CC therapy. The polynucleotide sequences encoding hGDMRP-1 may be used for
 CC diagnosing a disorder associated with the expression of hGDMRP-1, in
 CC particular heart and skeletal muscle disorders. hGDMRP-1 is localised to
 CC chromosome 22. The present sequence represents an oligomer used in the
 CC screening of the hGDMRP-1 sequence in the exemplification of the present
 CC invention.
 CC N.B. The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequence.

CC Sequence 25 BP; 6 A; 6 C; 11 G; 2 T; 0 other;

Query Match 76.7%; Score 13.8; DB 24; Length 25;
 Best Local Similarity 88.2%; Pred. No. 3e+03; Mismatches 2; Indels 0; Gaps 0;

Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 1 CTCAGAGAGAGAGAGC 17
 |||||
 6 CTCAGAGAGAGAGAGC 22

RESULT 9
 ID ABN05560

ABN05560 standard; DNA; 25 BP.

XX ABN05560;

XX 29-MAY-2002 (first entry)

DE Human GDMRP-1 25-mer scanning SEQ ID NO:4 sequence SEQ ID NO:5552.

KW Human; genome-derived myosin-like protein 1; GDMRP-1; hGDMRP-1; heart;
 KW muscle; myosin; chromosome 22; gene therapy; vaccine; heart disease;
 KW skeletal muscle disorder; amplicon; screening; ss.

OS Homo sapiens.

PN WO200192524-A2.

PD 06-DEC-2001.

PF 25-MAY-2001; 2001WO-US16981.

PR 26-MAY-2000; 2000US-207456P.

PR 21-SEP-2000; 2000US-234687P.

PR 27-SEP-2000; 2000US-236359P.

PR 04-OCT-2000; 2000GB-0024263.

PR 30-JAN-2001; 2001WO-US00661.

PR 30-JAN-2001; 2001WO-US00662.

PR 30-JAN-2001; 2001WO-US00663.

PR 30-JAN-2001; 2001WO-US00664.

PR 30-JAN-2001; 2001WO-US00665.

PR 30-JAN-2001; 2001WO-US00666.

PR 30-JAN-2001; 2001WO-US00667.

PR 30-JAN-2001; 2001WO-US00668.

PR 30-JAN-2001; 2001WO-US00669.

PR 05-FEB-2001; 2001US-266860P.

PI Gu Y, Ji Y, Penn SG, Hanzel DK, Rank DR, Chen W, Shannon WE;
 XX WPI; 2002-179446/23.
 DR New polypeptide, for raising antibodies that recognize hGDMRP-1
 XX proteins, or as specific biomolecule capture probes for
 PT surface-enhanced laser desorption ionization, comprises human
 PT myosin-like protein hGDMRP-1.
 PS Disclosure; SEQ ID 5552; 214p; English.
 XX The present invention describes a human genome-derived myosin-like
 CC protein 1 (hGDMRP-1). The protein and polynucleotide sequences of
 CC hGDMRP-1 can be used in gene therapy and vaccine production. The
 CC hGDMRP-1 nucleic acids can be used as probes to detect, characterise
 CC and quantify hGDMRP-1 nucleic acids in samples, as amplification
 CC substrates, to provide initial substrates for the recombinant engineering
 CC of hGDMRP-1 protein variants having desired phenotypic improvements, and
 CC for expressing the proteins. The hGDMRP-1 proteins or polypeptides may
 CC be used as immunogens to raise antibodies that specifically recognise
 CC hGDMRP-1 proteins, as standards in assays used to determine the
 CC concentration and/or amount specifically of hGDMRP proteins, as specific
 CC biomolecule capture probes for surface-enhanced laser desorption
 CC ionisation, as therapeutic supplement in patients having specific
 CC deficiency in hGDMRP-1 production, and in vaccines or for replacement
 CC therapy. The polynucleotide sequences encoding hGDMRP-1 may be used for
 CC diagnosing a disorder associated with the expression of hGDMRP-1, in
 CC particular heart and skeletal muscle disorders. hGDMRP-1 is localised to
 CC chromosome 22. The present sequence represents an oligomer used in the
 CC screening of the hGDMRP-1 sequence in the exemplification of the present
 CC invention.
 CC N.B. The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequence.

Sequence 25 BP; 6 A; 7 C; 10 G; 2 T; 0 other;

Query Match 76.7%; Score 13.8; DB 24; Length 25;
 Best Local Similarity 88.2%; Pred. No. 3e+03; Mismatches 2; Indels 0; Gaps 0;

Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 1 CTCAGAGAGAGAGAGC 17
 |||||
 5 CTCAGAGAGAGAGAGC 21

RESULT 10
 ID ABN05561

ABN05561 standard; DNA; 25 BP.

XX ABN05561;

XX 29-MAY-2002 (first entry)

DE Human GDMRP-1 25-mer scanning SEQ ID NO:4 sequence SEQ ID NO:5553.

KW Human; genome-derived myosin-like protein 1; GDMRP-1; hGDMRP-1; heart;
 KW muscle; myosin; chromosome 22; gene therapy; vaccine; heart disease;
 KW skeletal muscle disorder; amplicon; screening; ss.

OS Homo sapiens.

PN WO200192524-A2.

PD 06-DEC-2001.

PF 25-MAY-2001; 2001WO-US16981.

PR 26-MAY-2000; 2000US-207456P.

PR 21-SEP-2000; 2000US-234687P.

PR 27-SEP-2000; 2000US-236359P.

PR 04-OCT-2000; 2000GB-0024263.

PR 30-JAN-2001; 2001WO-US00661.

PR 30-JAN-2001; 2001WO-US00662.

PR 30-JAN-2001; 2001WO-US00663.

PR 30-JAN-2001; 2001WO-US00664.

PR 30-JAN-2001; 2001WO-US00665.

PR 30-JAN-2001; 2001WO-US00666.

PR 30-JAN-2001; 2001WO-US00667.

PR 30-JAN-2001; 2001WO-US00668.

PR 30-JAN-2001; 2001WO-US00669.

PR 05-FEB-2001; 2001US-266860P.

(AEOM -) AEOMICA INC.

PR 30-JAN-2001; 2001WO-US00662.
PR 30-JAN-2001; 2001WO-US00663.
PR 30-JAN-2001; 2001WO-US00664.
PR 30-JAN-2001; 2001WO-US00665.
PR 30-JAN-2001; 2001WO-US00666.
PR 30-JAN-2001; 2001WO-US00667.
PR 30-JAN-2001; 2001WO-US00668.
PR 30-JAN-2001; 2001WO-US00669.
PR 30-JAN-2001; 2001WO-US00670.
PR 05-FEB-2001; 2001US-266860P.
XX
XX
XX (AEOM-) AEOMICA INC.
XX
XX
XX Gu Y, Ji Y, Penn SG, Hanzel DK, Rank DR, Chen W, Shannon ME;
XX
XX WPI; 2002-179446/23.
XX
XX
XX New polypeptide, for raising antibodies that recognise hcdMLP-1
XX proteins, or as specific biomolecule capture probes for
XX surface-enhanced laser desorption/ionization, comprises human
XX myosin-like protein hcdMLP-1 -
XX
XX Disclosure; SEQ ID 5553; 214pp; English.

CC	The present invention describes a human genome-derived myosin-like
CC	protein 1 (hgDMLP-1). The protein and polynucleotide sequences of
CC	hgDMLP-1 can be used in gene therapy and vaccine production. The
CC	hgDMLP-1 nucleic acids can be used as probes to detect, characterise
CC	and quantify hgDMLP-1 nucleic acids in samples, as amplification
CC	substrates, to provide initial substrates for the recombinant engineering
CC	of hgDMLP-1 protein variants having desired phenotypic improvements, and
CC	for expressing the proteins. The hgDMLP-1 proteins or polypeptides may
CC	be used as immunogens to raise antibodies that specifically recognise
CC	hgDMLP-1 proteins, as standards in assays used to determine the
CC	concentration and/or amount specifically of hgDMLP proteins, as specific
CC	biomolecule capture probes for surface-enhanced laser desorption
CC	ionisation, as therapeutic supplement in patients having specific
CC	deficiency in hgDMLP-1 production, and in vaccines or for replacement
CC	therapy. The polynucleotide sequences encoding hgDMLP-1 may be used for
CC	diagnosing a disorder associated with the expression of hgDMLP-1, in
CC	particular heart and skeletal muscle disorders. hgDMLP-1 is localised to
CC	chromosome 22. The present sequence represents an oligomer used in the
CC	screening of the hgDMLP-1 sequence in the exemplification of the present
CC	invention.
CC	N.B. The sequence data for this patent did not form part of the printed
CC	specification, but was obtained in electronic format directly from WIPO
CC	at ftp.wipo.int/pub/published_pct_sequence.
XX	
SQ	Sequence 25 BP; 6 A; 7 C; 9 G; 3 T; 0 other;
	Query Match 76.7%; Score 13.8; DB 24; Length 25;
	Best Local Similarity 88.2%; Pred. No. 3e+03; 2;
Matches	15; Conservative 0; Mismatches 2; Indels 0; Gaps 0
DY	1 CTCAGAGAGAGAGAGC 17
Db	4 CTCAGAGAGCGAGAGC 20
RESULT 11	
ABN05562	
ID	ABN05562 standard; DNA: 25 BP.
XX	
AC	ABN05562;
XX	
DT	29-MAY-2002 (first entry)
DE	
XX	
HM	Human GDMLP-1 25-mer scanning SEQ ID NO:4 sequence SEQ ID NO:5554.
KW	Human; genome-derived myosin-like protein 1; GDMLP-1; heart;
KW	muscle; myosin; chromosome 22; gene therapy; vaccine; heart disease;
KW	skeletal muscle disorder; amplicon; screening; ss.
XX	

OS	Homo sapiens.
XX	
PN	WO200192524-A2.
XX	
PD	06-DEC-2001.
XX	
XX	
PF	25-MAY-2001; 2001WO-US16981.
XX	
PR	26-MAY-2000; 2000US-207456P.
PR	21-SEP-2000; 2000US-234687P.
PR	27-SEP-2000; 2000US-236359P.
PR	04-OCT-2000; 2000GB-0024263.
PR	30-JAN-2001; 2001WO-US00661.
PR	30-JAN-2001; 2001WO-US00662.
PR	30-JAN-2001; 2001WO-US00663.
PR	30-JAN-2001; 2001WO-US00664.
PR	30-JAN-2001; 2001WO-US00665.
PR	30-JAN-2001; 2001WO-US00666.
PR	30-JAN-2001; 2001WO-US00667.
PR	30-JAN-2001; 2001WO-US00668.
PR	30-JAN-2001; 2001WO-US00669.
PR	30-JAN-2001; 2001WO-US00670.
PR	05-FEB-2001; 2001US-266860P.
XX	
PA	(AEMOM-) AEMOMICA INC.
PI	Gu Y, Ji Y, Penn SG, Hanzel DK, Rank DR, Chen W, Shannon ME.
XX	
DR	WPI; 2002-179446/23.

PT New polypeptide, for raising antibodies that recognize hGDMLP-1
 PT proteins, or as specific biomolecule capture probes for
 PT surface-enhanced laser desorption/ionization, comprises human
 XX myosin-like protein hGDMLP-1 -
 XX
 PS Disclosure; SEQ ID 5554; 214pp; English.
 XX
 CC The present invention describes a human genome-derived myosin-like
 CC protein 1 (hGDMLP-1). The protein and polynucleotide sequences of
 CC hGDMLP-1 can be used in gene therapy and vaccine production. The
 CC hGDMLP-1 nucleic acids can be used as probes to detect, characterise
 CC and quantify hGDMLP-1 nucleic acids in samples, as amplification
 CC substrates, to provide initial substrates for the recombinant engineering
 CC of hGDMLP-1 protein variants having desired phenotypic improvements, and
 CC for expressing the proteins. The hGDMLP-1 proteins or polypeptides may
 CC be used as immunogens to raise antibodies that specifically recognise
 CC hGDMLP-1 proteins, as standards in assays used to determine the
 CC concentration and/or amount specifically of hGDMLP proteins, as specific
 CC biomolecule capture probes for surface-enhanced laser desorption
 CC ionisation, as therapeutic supplement in patients having specific
 CC deficiency in hGDMLP-1 production, and in vaccines or for replacement
 CC therapy. The polynucleotide sequences encoding hGDMLP-1 may be used for
 CC diagnosing a disorder associated with the expression of hGDMLP-1, in
 CC particular heart and skeletal muscle disorders. hGDMLP-1 is localised to
 CC chromosome 22. The present sequence represents an amplification of the
 CC screening of the hGDMLP-1 sequence in the exemplification of the present
 CC invention.
 CC N.B. The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequence.
 XX
 SEQ Sequence 25 BP; 6 A; 6 C; 10 G; 3 T; 0 other;

```

Query Match          76.7%; Score 13.8; DB 24; Length 25;
Best Local Similarity 88.2%; Pred. No. 3e+03;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 CTCAGAGAGAGAGAGC 17
          |||||  |||  |||
DB      3 CTCAGAGAGCGAGCAGC 19

RESUM.T 12

```

ABN05563	ABN05563 standard; DNA: 25 BP.	
AC	ABN05563;	
AD		
AE	29-MAY-2002 (first entry)	
AF		
AG	Human GDMLP-1 25-mer scanning SEQ ID NO:4 sequence SEQ ID NO:5555.	
AH		
AI		
AJ	Human; genome-derived myosin-like protein 1; GDMLP-1; hGDMLP-1; heart; muscle; myosin; chromosome 22; gene therapy; vaccine; heart disease; skeletal muscle disorder; amplicon; screening; ss.	
AK		
AL	Homo sapiens.	
AM		
AN	WO200192524-A2.	
AO		
AP	06-DEC-2001.	
AQ		
AR	25-MAY-2001; 2001WO-US16981.	
AS		
AT	26-MAY-2000; 2000US-207456P.	
AV	21-SEP-2000; 2000US-234687P.	
AW	27-SEP-2000; 2000US-236359P.	
AX	04-OCT-2000; 2000GB-0024263.	
AY	30-JAN-2001; 2001WO-US00661.	
AZ	30-JAN-2001; 2001WO-US00662.	
BA	30-JAN-2001; 2001WO-US00663.	
BB	30-JAN-2001; 2001WO-US00664.	
BC	30-JAN-2001; 2001WO-US00665.	
BD	30-JAN-2001; 2001WO-US00666.	
BE	30-JAN-2001; 2001WO-US00667.	
BF	30-JAN-2001; 2001WO-US00668.	
BG	30-JAN-2001; 2001WO-US00669.	
BH	30-JAN-2001; 2001WO-US00670.	
BI	05-FEB-2001; 2001US-266860P.	
BJ		
BK	(ABEOM-) ABEOMICA INC.	
BL		
BM	Gu Y, Ji Y, Penn SG, Hanzel DK, Rank DR, Chen W, Shannon ME;	
BN	WPI: 2002-179446/23.	
BO		
BP		
BQ		
BR		
BS		
BT		
BU		
BV		
BW		
BX		
BY		
BZ		
CA		
CB		
CC		
CD		
CE		
CF		
CG		
CH		
CI		
CJ		
CK		
CL		
CM		
CN		
CO		
CP		
CQ		
CR		
CS		
CT		
CU		
CV		
CW		
CX		
CY		
CA		
CB		
CC		
CD		
CE		
CF		
CG		
CH		
CI		
CJ		
CK		
CL		
CM		
CN		
CO		
CP		
CQ		
CR		
CS		
CT		
CU		
CV		
CW		
CX		
CY		
CA		
CB		
CC		
CD		
CE		
CF		
CG		
CH		
CI		
CJ		
CK		
CL		
CM		
CN		
CO		
CP		
CQ		
CR		
CS		
CT		
CU		
CV		
CW		
CX		
CY		
CA		
CB		
CC		
CD		
CE		
CF		
CG		
CH		
CI		
CJ		
CK		
CL		
CM		
CN		
CO		
CP		
CQ		
CR		
CS		
CT		
CU		
CV		
CW		
CX		
CY		
CA		
CB		
CC		
CD		
CE		
CF		
CG		
CH		
CI		
CJ		
CK		
CL		
CM		
CN		
CO		
CP		
CQ		
CR		
CS		
CT		
CU		
CV		
CW		
CX		
CY		
CA		

SQ	Sequence	25 BP; 5 A; 6 C; 10 G; 4 T; 0 other;
Query Match		76.7%; Score 13.8; DB 24; Length 25;
Best Local Similarity		88.2%; Pred. No. 3e+03;
Matches	15; Conservative	0; Mismatches 2; Indels 0; Gaps 0;
OY	1 CTCAGAGAGAGAGAGC 17 Db 2 CTCAGAGAGCGACGAC 18	
RESULT 13		
ABN05564		
ID	ABN05564 standard; DNA; 25 BP.	
XX AC	ABN05564;	
XX DT	29-MAY-2002 (first entry)	
DE	Human GDMLP-1 25-mer scanning SEQ ID NO:4 sequence SEQ ID NO:5556.	
XX KW	Human; genome-derived myosin-like protein 1; GDMLP-1; hGDMLP-1; heart; muscle; myosin; chromosome 22; gene therapy; vaccine; heart disease; skeletal muscle disorder; amplicon; screening; ss.	
XX OS	Homo sapiens.	
XX PN	WO200192524-A2.	
XX PD	06-DEC-2001.	
XX PE	25-MAY-2001; 2001WO-US16981.	
XX PR	26-MAY-2000; 2000US-207456P. 21-SEP-2000; 2000US-234687P. 27-SEP-2000; 2000US-236359P. 04-OCT-2000; 2000GB-0024263. 30-JAN-2001; 2001WO-US00661. 30-JAN-2001; 2001WO-US00662. 30-JAN-2001; 2001WO-US00663. 30-JAN-2001; 2001WO-US00664. 30-JAN-2001; 2001WO-US00665. 30-JAN-2001; 2001WO-US00666. 30-JAN-2001; 2001WO-US00667. 30-JAN-2001; 2001WO-US00668. 30-JAN-2001; 2001WO-US00669. 30-JAN-2001; 2001WO-US00670. 05-FEB-2001; 2001US-266860P.	
XX PA	(AEOM-) AEOMICA INC.	
XX PI	Gu Y, Ji Y, Penn SG, Hanzel DK, Rank DR, Chen W, Shannon ME;	
DR	WPI; 2002-179446/23.	
XX PT	New polypeptide, for raising antibodies that recognize hGDMLP-1 proteins, or as specific biomolecule capture probes for surface-enhanced laser desorption/ionization, comprises human myosin-like protein hGDMLP-1 -	
XX PS	Disclosure; SEQ ID 5556; 214pp; English.	
XX CC	The present invention describes a human genome-derived myosin-like protein 1 (hGDMLP-1). The protein and polynucleotide sequences of hGDMLP-1 can be used in gene therapy and vaccine production. The hGDMLP-1 nucleic acids can be used as probes to detect, characterise and quantify hGDMLP-1 nucleic acids in samples, as amplification substrates, to provide initial substrates for the recombinant engineering of hGDMLP-1 protein variants having desired phenotypic improvements, and for expressing the proteins. The hGDMLP-1 proteins or polypeptides may be used as immunogens to raise antibodies that specifically recognise hGDMLP-1 proteins, as standards in assays used to determine the	

CC concentration and/or amount specifically of hGDMRP proteins, as specific
 CC biomolecule capture probes for surface-enhanced laser desorption
 CC ionisation, as therapeutic supplement in patients having specific
 CC deficiency in hGDMRP-1 production, and in vaccines or for replacement
 CC therapy. The polynucleotide sequences encoding hGDMRP-1 may be used for
 CC diagnosing a disorder associated with the expression of hGDMRP-1, in
 CC particular heart and skeletal muscle disorders. hGDMRP-1 is localised to
 CC chromosome 22. The present sequence represents an oligomer used in the
 CC screening of the hGDMRP-1 sequence in the exemplification of the present
 CC invention.
 CC N.B. The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at [ftp.wipo.int/pub/published_pct_sequence](http://wipo.int/pub/published_pct_sequence).
 CC
 CC Sequence 25 BP; 5 A; 6 C; 10 G; 4 T; 0 other;
 SQ
 Query Match 76.7%; Score 13.8; DB 24; Length 25;
 Best Local Similarity 88.2%; Pred. No. 3e+03; 2; Indels 0; Gaps 0;
 Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 CTCAGAGAGAGAGAGC 17
 1 CTCAGAGAGAGAGAGC 17
 Db 1 CTCAGAGAGAGAGAGC 17
 RESULT 14
 AAD39624/C
 ID AAD39624 standard; DNA; 20 BP.
 XX
 AC AAD39624;
 XX
 XX 04-OCT-2002 (first entry)
 XX
 DE Human SR-cyp antisense oligonucleotide, ISIS #123888.
 XX
 KW Human; antisense; SR-cyp; CLK-associated RS cyclophilin; inflammation;
 KM hyperproliferative disorder; cancer; prophylaxis; infection; therapy;
 KW tumour; CARs-cyp; phosphorothioate backbone; ss.
 XX
 OS Homo sapiens.
 OS Synthetic.
 XX
 FH key
 FT modified_base
 FT 1..20
 FT /tag- a
 FT /mod_base- OTHER
 FT /note- "2-methoxyethyl nucleotides"
 FT 16..20
 FT /tag- c
 FT /mod_base- OTHER
 FT /note- "2-methoxyethyl nucleotides"
 FT 2
 FT /tag- d
 FT /mod_base- m5c
 FT 5
 FT /tag- e
 FT /mod_base- m5c
 FT 7
 FT /tag- f
 FT /mod_base- m5c
 FT 9
 FT /tag- g
 FT /mod_base- m5c
 FT 11
 FT /tag- h
 FT /mod_base- m5c
 FT 16
 FT /tag- i
 FT /mod_base- m5c
 FT modified_base

FT modified_base 18..19
 FT /tag- j
 FT /mod_base- m5c
 XX
 XX WO200236809-A2.
 XX
 XX 10-MAY-2002.
 PD
 PD 30-OCT-2001; 2001WO-US47335.
 PF
 XX
 XX 03-NOV-2000; 2000US-0706197.
 XX
 XX (ISIS-) ISIS PHARM INC.
 PA (COLD-) COLD SPRING HARBOR LAB.
 PA
 PI Bennett CF, Spector DL, Wyatt JR;
 XX
 DR WPI; 2002-479763/51.
 XX
 XX Novel antisense compounds targeted to nucleic acids encoding SR-cyp,
 PT CLK-associated RS cyclophilin for modulating the gene expression and
 PT treating hyperproliferative disorders such as cancer
 XX
 XX Claim 3; Page 90; 117pp; English.
 PS
 XX The invention relates to antisense compounds targeted to a nucleic
 CC acid molecule encoding human SR-cyp (CLK-associated RS cyclophilin)
 CC to inhibit its expression. SR-cyp is also referred to as CARs-cyp.
 CC Antisense compounds of the invention are used for treating diseases
 CC or conditions associated with SR-cyp. The diseases treated include
 CC hyperproliferative disorders e.g. cancer or hyperproliferative
 CC disorders resulting from an alternative splicing event. They are
 CC useful for diagnostics, therapeutics and as research reagents, e.g.
 CC prophylactically to prevent or delay infection, inflammation or
 CC tumour formation. They are also used in antisense therapy. The present
 CC sequence is an antisense oligonucleotide targeted to human SR-cyp.
 XX
 SQ Sequence 20 BP; 1 A; 8 C; 3 G; 8 T; 0 other;
 Query Match 74.4%; Score 13.4; DB 24; Length 20;
 Best Local Similarity 93.3%; Pred. No. 4.4e+03;
 Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 CTCAGAGAGAGAGA 15
 15 CTCAGAGAGAGAGA 1
 Db 15 CTCAGAGAGAGAGA 1
 RESULT 15
 AAH62255
 ID AAH62255 standard; DNA; 21 BP.
 XX
 AC AAH62255;
 XX
 XX 12-SEP-2001 (first entry)
 DE
 DE SLC1A6 polymorphism containing DNA fragment #156.
 XX
 KW Single nucleotide polymorphism; SNP; human; cancer; inflammation;
 KM heart disease; paternity testing; forensic science; ds.
 KW
 XX Homo sapiens.
 OS
 OS
 FH key
 FT variation
 FT location/qualifiers
 FT replace(11,G)
 FT /tag- a
 FT /standard_name- "single nucleotide polymorphism"
 FT WO200138576-A2.
 XX
 XX 31-MAY-2001.
 PD
 PD 17-NOV-2000; 2000WO-US31639.
 PF

XX 24-NOV-1999; 99US-0167334.
PR
XX
PA (WHED) WHITEHEAD INST BIOMEDICAL RES.
XX
XX
PI Cargill M, Ireland JS, Lander ES;
XX
DR WPI; 2001-367705/38.
XX
XX
PT New nucleic acid segments of the human genome, particularly from genes
PT including polymorphic sites, for phenotype correlation, forensics,
PT paternity testing, medicine and genetic analysis -
XX
PS Claim 1; Page 42; 80pp; English.
XX
CC DNA sequences AAH62100 - AAH62688 represent segments of human genes which
CC contain single nucleotide polymorphisms (SNPs). A method is included in
CC the invention for analysing a nucleic acid sample, which consists of
CC determining the base occupying any one of the polymorphic sites given in
CC the SNP containing sequences. The nucleotide sequences can be used in the
CC diagnosis or monitoring of diseases, such as cancer, inflammation, heart
CC diseases, diseases of the cardiovascular system, and infection by
CC microorganisms. The oligonucleotides are also useful in the manufacture
CC of a medicament for the treatment or prophylaxis of the diseases, and as
CC a pharmaceutical. SNP containing oligonucleotides are useful in
CC applications such as phenotype correlation, forensics, paternity testing,
CC medicine and genetic analysis.
XX
SQ Sequence 21 BP; 6 A; 5 C; 8 G; 2 T; 0 other;
Query Match 74.4%; Score 13.4; DB 22; Length 21;
Best Local Similarity 93.3%; Pred. No. 4.4e+03;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 CTCAGAGAGAGAGCA 15
| | | | | | | | | | | | | | | | | | | | | |
Db 4 CTCAGAGAGAGAGCA 18

Search completed: November 9, 2002, 14:12:16
Job time : 297 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 9, 2002, 14:05:36 ; Search time 67 Seconds
(without alignments)
82.391 Million cell updates/sec

Title: US-09-902-772-1_COPY_645_662
Perfect score: 18
Sequence: 1 ctaagagagagagagacc 18

Scoring table:
IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 687286

Minimum DB seq length: 0
Maximum DB seq length: 100 \

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /cgn2_6/ptodata/1/lna/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/lna/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/lna/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/lna/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/lna/PCMTUS.COMB.seq:*
6: /cgn2_6/ptodata/1/lna/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Query Length	DB ID	Description
1	13.4	74.4	27	1 US-08-004-552-2	Sequence 2, Appl1
2	13	72.2	22	1 US-08-592-126-24	Sequence 24, Appl1
3	13	72.2	22	1 US-08-687-080-24	Sequence 24, Appl1
4	13	72.2	23	1 US-07-946-239-5	Sequence 5, Appl1
5	13	72.2	23	1 US-09-151-467-5	Sequence 5, Appl1
6	13	72.2	23	3 US-09-036-599-10	Sequence 10, Appl1
7	13	72.2	23	4 US-09-256-838-5	Sequence 5, Appl1
8	13	72.2	23	5 PCT-US92-07815-5	Sequence 5, Appl1
9	13	72.2	36	1 US-08-863-639A-26	Sequence 26, Appl1
10	13	72.2	43	1 US-08-560-113A-8	Sequence 8, Appl1
11	13	72.2	43	1 US-08-611-155B-12	Sequence 12, Appl1
12	13	72.2	43	2 US-08-916-120A-14	Sequence 14, Appl1
13	13	72.2	18	3 US-08-389-360-6	Sequence 6, Appl1
14	13	72.2	18	3 US-09-161-344-26	Sequence 26, Appl1
15	13	72.2	18	3 US-09-038-328-6	Sequence 6, Appl1
16	13	72.2	18	4 US-09-183-931-15	Sequence 15, Appl1
17	13	72.2	18	4 US-09-705-160-15	Sequence 15, Appl1
18	13	72.2	18	4 US-09-435-524-6	Sequence 15, Appl1
19	13	72.2	18	4 US-08-646-367-19	Sequence 19, Appl1
20	13	72.2	33	2 US-08-275-850-255	Sequence 255, App
21	12.4	68.9	20	1 US-08-164-200-5	Sequence 3, Appl1
22	12.4	68.9	35	4 US-09-450-072-37	Sequence 37, Appl1
23	12.4	68.9	35	4 US-09-351-348-37	Sequence 37, Appl1
24	12.4	68.9	41	5 PCT-US92-10792-28	Sequence 28, Appl1
25	12.4	68.9	45	1 US-09-640-304-13	Sequence 13, Appl1
26	12.4	68.9	45	1 US-08-190-103-13	Sequence 13, Appl1
27	12.4	68.9	45	1 US-08-454-778-5	Sequence 5, Appl1

28	12.4	68.9	47	4 US-09-455-960-14	Sequence 14, Appl1
29	12.4	68.9	50	2 US-09-200-232-4	Sequence 4, Appl1
30	12.4	68.9	69	2 US-08-523-125-7	Sequence 7, Appl1
31	12.4	68.9	69	3 US-08-660-561A-7	Sequence 7, Appl1
32	12.4	68.9	87	2 US-08-523-125-10	Sequence 10, Appl1
33	12.4	68.9	87	3 US-08-660-561A-10	Sequence 10, Appl1
34	12.4	68.9	87	3 PCT-US92-10792-23	Sequence 23, Appl1
35	12.4	68.9	87	5 US-08-314-596-34	Sequence 34, Appl1
36	12.4	68.9	90	1 US-08-320-982-34	Sequence 34, Appl1
37	12.4	68.9	90	3 US-08-819-037-34	Sequence 34, Appl1
38	12.4	68.9	90	4 US-09-045-940-34	Sequence 34, Appl1
39	12.4	68.9	100	1 US-08-129-930B-90	Sequence 9, Appl1
40	12.4	68.9	100	4 US-08-134-346A-45	Sequence 45, Appl1
41	12.4	68.9	100	4 US-08-976-288A-90	Sequence 90, Appl1
42	12.4	68.9	100	4 US-08-377-503-5	Sequence 5, Appl1
43	12.4	68.9	100	4 US-08-377-503-5	Sequence 5, Appl1
44	12.4	68.9	100	4 US-08-178-019-5	Sequence 5, Appl1
45	12.4	68.9	100	4 US-08-178-019-5	Sequence 5, Appl1

ALIGNMENTS

RESULT 1
US-08-004-552-2/c
Sequence 2, Application US/08004552
Patent No. 5482836
GENERAL INFORMATION:
APPLICANT: Cantor, Charles R.
APPLICANT: Ito, Takashi
TITLE OF INVENTION: DNA PURIFICATION BY TRIPLEX-AFFINITY
TITLE OF INVENTION: CAPTURE AND AFFINITY CAPTURE ELECTROPHORESIS
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Karen S. Smith
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: CA 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/004.552
FILING DATE: 19930114
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Smith, Karen S.
REGISTRATION NUMBER: 31,426
REFERENCE/DOCKET NUMBER: A-57666/KSS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 27 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cdna
FEATURE:
NAME/KEY: misc-feature
LOCATION: 1..27
OTHER INFORMATION: /note="BamTC oligonucleotide."
US-08-004-552-2
Query Match 74.4%; Score 13.4; DB 1; Length 27;
Best Local Similarity 93.3%; Pred. No. 5.2e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 4 AGAGAGAGAGAGCC 18
Db 18 AGAGAGAGAGATCC 4

RESULT 2
US-08-592-126-24/c
; Sequence 24, Application US/08592126
; Patent No. 5821091

GENERAL INFORMATION:

APPLICANT: Gregory Dolganov

TITLE OF INVENTION: Transcripts Encoding Immunomodulatory

NUMBER OF SEQUENCES: 151

CORRESPONDENCE ADDRESS:

ADDRESSEE: Dehlinger & Associates

STREET: 350 Cambridge Avenue, Suite 250

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94306

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/592,126

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Sholtz, Charles K.

REGISTRATION NUMBER: 38,615

REFERENCE/DOCKET NUMBER: 4600-0111

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 324-0880

TELEFAX: (415) 324-0960

INFORMATION FOR SEQ ID NO: 24:

SEQUENCE CHARACTERISTICS:

LENGTH: 22 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

MOLECULE TYPE: DNA

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

INDIVIDUAL ISOLATE: Primer EGR11-6

US-08-592-126-24

Query Match 72.2%; Score 13; DB 1; Length 22;
Best Local Similarity 100.0%; Pred. No. 7.8e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 AGAGAGAGAGAG 16
Db 17 AGAGAGAGAGAG 5

RESULT 3
US-08-687-080-24/c

; Sequence 24, Application US/08687080

; Patent No. 5965427

GENERAL INFORMATION:

APPLICANT: Gregory Dolganov

TITLE OF INVENTION: Human RAD50 Gene and Methods of Use Thereof

NUMBER OF SEQUENCES: 175

CORRESPONDENCE ADDRESS:

ADDRESSEE: Dehlinger & Associates

STREET: 350 Cambridge Avenue, Suite 250

CITY: Palo Alto

STATE: CA

COUNTRY: USA
ZIP: 94306

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/687,080

FILING DATE: 17-JUL-1996

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/592,126

FILING DATE: 26-JAN-1996

ATTORNEY/AGENT INFORMATION:

NAME: Sholtz, Charles K.

REGISTRATION NUMBER: 38,615

REFERENCE/DOCKET NUMBER: 4600-0111.30

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 324-0880

TELEFAX: (415) 324-0960

INFORMATION FOR SEQ ID NO: 24:

SEQUENCE CHARACTERISTICS:

LENGTH: 22 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

MOLECULE TYPE: linear

TOPOLOGY: linear

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

INDIVIDUAL ISOLATE: Primer EGR11-6

US-08-687-080-24

Query Match 72.2%; Score 13; DB 2; Length 22;
Best Local Similarity 100.0%; Pred. No. 7.8e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 AGAGAGAGAGAG 16
Db 17 AGAGAGAGAGAG 5

RESULT 4
US-07-946-239-5

; Sequence 5, Application US/07946239

; Patent No. 5770358

GENERAL INFORMATION:

APPLICANT: DOWER, WILLIAM J

APPLICANT: BARRETT, RONALD W

APPLICANT: GALLOP, MARK A

APPLICANT: NEEDLES, MICHAEL C

TITLE OF INVENTION: METHOD OF SYNTHESIZING DIVERSE

NUMBER OF SEQUENCES: 16

CORRESPONDENCE ADDRESS:

ADDRESSEE: TOWNSEND AND TOWNSEND

STREET: 1 MARKET PLAZA, STEWART TOWER, SUITE 2000

CITY: SAN FRANCISCO

STATE: CALIFORNIA

COUNTRY: USA

ZIP: 94105

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/946,239

FILING DATE: 19920916

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Smith, William M.

```

:      REGISTRATION NUMBER: 30,223
:      REFERENCE/DOCKET NUMBER: 11509-36-1
:      TELECOMMUNICATION INFORMATION:
:      TELEPHONE: 415-543-9600
:      TELEFAX: 415-543-5043
:      INFORMATION FOR SEQ ID NO: 5:
:      SEQUENCE CHARACTERISTICS:
:      LENGTH: 23 base pairs
:      TYPE: NUCLEIC ACID
:      STRANDEDNESS: single
:      TOPOLOGY: linear
:      MOLECULE TYPE: cDNA
:      US-07-946-239-5

Query Match      72.2%; Score 13; DB 1; Length 23;
Best Local Similarity 100.0%; Pred. No. 7.8e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      4 AGAGAGAGAGAG 16
      |||||||
Db      8 AGAGAGAGAGAG 20

RESULT 5
US-09-151-467-5
: Sequence 5, Application US/09151467
: Patent No. 6140493
: GENERAL INFORMATION:
: APPLICANT: DOWER, WILLIAM J
: APPLICANT: BARRETT, RONALD W
: APPLICANT: GALLOP, MARK A
: APPLICANT: NEEDELS, MICHAEL C
: TITLE OF INVENTION: METHOD OF SYNTHESIZING DIVERSE
: TITLE OF INVENTION: COLLECTIONS OF OLIGOMERS
: NUMBER OF SEQUENCES: 16
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: TOWNSEND AND TOWNSEND
: STREET: 1 MARKET PLAZA, STEUART TOWER, SUITE 2000
: CITY: SAN FRANCISCO
: STATE: CALIFORNIA
: COUNTRY: USA
: ZIP: 94105
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/151,467
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/07/946,239
: FILING DATE: 1992-09-16
: ATTORNEY/AGENT INFORMATION:
: NAME: Smith, William M.
: REGISTRATION NUMBER: 30,223
: REFERENCE/DOCKET NUMBER: 11509-36-1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415-543-9600
: TELEFAX: 415-543-5043
: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 23 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: US-09-151-467-5

Query Match      72.2%; Score 13; DB 3; Length 23;
Best Local Similarity 100.0%; Pred. No. 7.8e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

QY      4 AGAGAGAGAGAG 16
      |||||||
Db      8 AGAGAGAGAGAG 20

RESULT 6
US-09-036-599-10
: Sequence 10, Application US/09036599
: Patent No. 6143497
: GENERAL INFORMATION:
: APPLICANT: Dower, William J.
: APPLICANT: Barrett, Ronald W.
: APPLICANT: Gallop, Mark A.
: TITLE OF INVENTION: Method of Synthesizing Diverse
: TITLE OF INVENTION: Collections of Oligomers
: NUMBER OF SEQUENCES: 12
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: William M. Smith
: STREET: One Market Plaza, Steuart Tower, Suite 2000
: CITY: San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94105
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/036,599
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/07/762,522
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Smith, William M.
: REGISTRATION NUMBER: 30,223
: REFERENCE/DOCKET NUMBER: 11509-36
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415-326-2400
: TELEFAX: 415-543-5043
: INFORMATION FOR SEQ ID NO: 10:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 23 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: US-09-036-599-10

Query Match      72.2%; Score 13; DB 3; Length 23;
Best Local Similarity 100.0%; Pred. No. 7.8e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      4 AGAGAGAGAGAG 16
      |||||||
Db      8 AGAGAGAGAGAG 20

RESULT 7
US-09-256-838-5
: Sequence 5, Application US/09256838
: Patent No. 6416949
: GENERAL INFORMATION:
: APPLICANT: DOWER, WILLIAM J
: APPLICANT: BARRETT, RONALD W
: APPLICANT: GALLOP, MARK A
: APPLICANT: NEEDELS, MICHAEL C
: TITLE OF INVENTION: METHOD OF SYNTHESIZING DIVERSE
: TITLE OF INVENTION: COLLECTIONS OF OLIGOMERS
: NUMBER OF SEQUENCES: 16
```

```

CORRESPONDENCE ADDRESS:
ADDRESS: TOWNSEND AND TOWNSEND
STREET: 1 MARKET PLAZA, STEWART TOWER, SUITE 2000
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94105

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09-256, 838
FILING DATE: 24-Feb-1999
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/946, 239
FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:
NAME: Smith, William M.
REGISTRATION NUMBER: 30, 223
REFERENCE/DOCKET NUMBER: 11509-36-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-543-9600
TELEFAX: 415-543-5043

INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 23 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-256-838-5

Query Match          72.2%; Score 13; DB 4; Length 23;
Best Local Similarity 100.0%; Pred. NO. 7.8e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      4 AGAGAGAGAGGAG 16
        |||
        8 AGAGAGAGAGGAG 20

RESULT 8
PCT-US92-07815-5
; Sequence 5, Application PC/TUS9207815
; GENERAL INFORMATION:
; APPLICANT: DOWER, WILLIAM J
; APPLICANT: BARRETT, RONALD W
; APPLICANT: GALLOP, MARK A
; APPLICANT: NEEDLES, MICHAEL C
; TITLE OF INVENTION: METHOD OF SYNTHESIZING DIVERSE
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND AND TOWNSEND
; STREET: 1 MARKET PLAZA, STEWART TOWER, SUITE 2000
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/07815
; FILING DATE: 19920916
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:

```

```

: NAME: Smith, William M.
: REGISTRATION NUMBER: 30,223
: REFERENCE/DOCKET NUMBER: 11509-36-1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415-543-9600
: TELEFAX: 415-543-5043
: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 23 base pairs
: TYPE: NUCLEIC ACID
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: CDNA
: PCT-US92-07815-5

Query Match          72.2%: Score 13; DB 5; Length 23;
Best Local Similarity 100.0%; Pred. No. 7.8e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      4 ACAGAGAGAGGAG 16
        |||||||
Db       8 ACAGAGAGAGGAG 20

RESULT 9
US-08-863-639A-26
: Sequence 26, Application US/08863639A
: Patent No. 5981185
: GENERAL INFORMATION:
: APPLICANT: Matson, Robert S.
: APPLICANT: Coassin, Peter J.
: APPLICANT: Rampal, Jang B.
: APPLICANT: Caskey, C. T.
: TITLE OF INVENTION: OLIGONUCLEOTIDE REPEAT ARRAYS
: NUMBER OF SEQUENCES: 95
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Sheldon & Mak
: STREET: 225 South Lake Avenue, 9th Floor
: CITY: Pasadena
: STATE: CA
: COUNTRY: USA
: ZIP: 91101
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage
: COMPUTER: IBM compatible
: OPERATING SYSTEM: Windows 95
: SOFTWARE: Corel Wordperfect 8 version
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/863,639A
: FILING DATE: May 28, 1997
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Joseph E. Mueeth
: REGISTRATION NUMBER: 20,532
: REFERENCE/DOCKET NUMBER: 11859-1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (626) 796-4000
: TELEFAX: (626) 795-6321
: INFORMATION FOR SEQ ID NO: 26:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 36 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: Other nucleic acid
: US-08-863-639A-26

Query Match          72.2%: Score 13; DB 2; Length 36;
Best Local Similarity 100.0%; Pred. No. 8e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      4 ACAGAGAGAGGAG 16
        |||||||

```

Db 16 AGAGAGAGAGAG 28

RESULT 10

US-08-560-313A-8

Sequence 8, Application US/08560313A

Patent No. 5763175

GENERAL INFORMATION:

APPLICANT: Sydney Brenner

TITLE OF INVENTION: Simultaneous Sequencing of Tagged Polynucleotides

NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:

ADDRESSEE: Stephen C. Macevitz, Lynx Therapeutics, Inc.

STREET: 3832 Bay Center Place

CITY: Hayward

STATE: California

COUNTRY: USA

ZIP: 94545

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 inch diskette

COMPUTER: Power Macintosh

OPERATING SYSTEM: Macintosh OS ver. 7.5.2

SOFTWARE: Microsoft Word, vers. 6.0.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/560,313A

FILING DATE: 17-NOV-95

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Stephen C. Macevitz

REGISTRATION NUMBER: 30,285

REFERENCE/DOCKET NUMBER: sstus

TELECOMMUNICATION INFORMATION:

TELEPHONE: (510) 670-9365

TELEFAX: (510) 670-9302

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:

LENGTH: 43 nucleotides

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

US-08-560-313A-8

Query Match 72.2%; Score 13; DB 1; Length 43;

Best Local Similarity 100.0%; Pred. No. 8.1e+02;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 AGAGAGAGAGAG 16

Db 11 AGAGAGAGAGAG 23

RESULT 11

US-08-611-155B-12

Sequence 12, Application US/08611155B

Patent No. 5780231

GENERAL INFORMATION:

APPLICANT: Sydney Brenner

TITLE OF INVENTION: DNA Extension and Analysis with Rolling Primers

NUMBER OF SEQUENCES: 13

CORRESPONDENCE ADDRESS:

ADDRESSEE: Stephen C. Macevitz, Lynx Therapeutics, Inc.

STREET: 3832 Bay Center Place

CITY: Hayward

STATE: California

COUNTRY: USA

ZIP: 94545

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 inch diskette

COMPUTER: IBM compatible

OPERATING SYSTEM: Windows 3.1

SOFTWARE: Microsoft Word, vers. 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/611,155B

FILING DATE: 05-MAR-96

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/560,313

FILING DATE: 17-NOV-95

ATTORNEY/AGENT INFORMATION:

NAME: Stephen C. Macevitz

REGISTRATION NUMBER: 30,285

REFERENCE/DOCKET NUMBER: srplus

TELECOMMUNICATION INFORMATION:

TELEPHONE: (510) 670-9365

TELEFAX: (510) 670-9302

INFORMATION FOR SEQ ID NO: 12:

SEQUENCE CHARACTERISTICS:

LENGTH: 43 nucleotides

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

US-08-611-155B-12

Query Match 72.2%; Score 13; DB 1; Length 43;

Best Local Similarity 100.0%; Pred. No. 8.1e+02;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 AGAGAGAGAGAG 16

Db 11 AGAGAGAGAGAG 23

RESULT 12

US-08-916-120A-14

Sequence 14, Application US/08916120A

Patent No. 5962228

GENERAL INFORMATION:

APPLICANT: Sydney Brenner

TITLE OF INVENTION: DNA Extension and Analysis with Rolling Primers

NUMBER OF SEQUENCES: 19

CORRESPONDENCE ADDRESS:

ADDRESSEE: Stephen C. Macevitz, Lynx Therapeutics, Inc.

STREET: 3832 Bay Center Place

CITY: Hayward

STATE: California

COUNTRY: USA

ZIP: 94545

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 inch diskette

COMPUTER: IBM compatible

OPERATING SYSTEM: Windows 95

SOFTWARE: Microsoft Word, vers. 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/916,120A

FILING DATE: 22-AUG-97

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/611,155

FILING DATE: 05-MAR-96

ATTORNEY/AGENT INFORMATION:

NAME: Stephen C. Macevitz

REGISTRATION NUMBER: 30,285

REFERENCE/DOCKET NUMBER: 811-01

TELECOMMUNICATION INFORMATION:

TELEPHONE: (510) 670-9365

TELEFAX: (510) 670-9302

INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:

LENGTH: 43 nucleotides

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

US-08-916-120A-14

Query Match 72.2%; Score 13; DB 2; Length 43;
Best Local Similarity 100.0%; Pred. No. 8.1e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 AGAGAGAGAGAG 16
DB 11 AGAGAGAGAGAG 23

RESULT 13
US-08-389-360-6/C

; Sequence 6, Application US/08389360
; Patent No. 5877017
; GENERAL INFORMATION:
; APPLICANT: van der Bruggen et al.
; TITLE OF INVENTION: ISOLATED PEPTIDE WHICH FORMS COMPLEXES
; TITLE OF INVENTION: WITH MHC MOLECULE HLA-C*1601 AND USBS
; NUMBER OF INVENTION: THEREOF
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022

; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch 1.44 Mb storage diskette

; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/389,360
; FILING DATE: Herewith

; CLASSIFICATION: 436
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/196,630
; FILING DATE: February 15, 1994

; APPLICATION DATA:
; APPLICATION NUMBER: 08/079,110
; FILING DATE: June 17, 1993

; ATTORNEY/AGENT INFORMATION:
; NAME: Pasqualini, Patricia A.
; REGISTRATION NUMBER: 34,894

; REFERENCE/DOCKET NUMBER: LUD 5310.2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200

; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 6:

; SEQUENCE CHARACTERISTICS:
; LENGTH: 18
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear

US-08-389-360-6

Query Match 71.1%; Score 12.8; DB 2; Length 18;
Best Local Similarity 87.5%; Pred. No. 9.6e+02;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 CAGAGAGAGAGGCC 18
DB 17 CAGAGAGAGAGGCC 2

RESULT 14
US-09-161-244-26/C

; Sequence 26, Application US/09161244
; Patent No. 6004814
; GENERAL INFORMATION:
; APPLICANT: Bennett, C. Frank
; APPLICANT: Cowsett, Lex M.

; TITLE OF INVENTION: ANTISENSE MODULATION OF CD71 EXPRESSION
; FILE REFERENCE: RTS-0007
; CURRENT APPLICATION NUMBER: US/09/161,244
; CURRENT FILING DATE: 1998-09-25
; NUMBER OF SEQ ID NOS: 91
; SEQ ID NO 26
; LENGTH: 18
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Antisense Oligonucleotide

US-09-161-244-26

Query Match 71.1%; Score 12.8; DB 3; Length 18;
Best Local Similarity 87.5%; Pred. No. 9.6e+02;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 CAGAGAGAGAGGCC 18
DB 18 CAGTGAAGAGAGGCC 3

RESULT 15

US-09-038-328-6/C
; Sequence 6, Application US/09038328
; Patent No. 6110694

; GENERAL INFORMATION:
; APPLICANT: van der Bruggen et al.

; TITLE OF INVENTION: ISOLATED PEPTIDE WHICH FORMS COMPLEXES
; TITLE OF INVENTION: WITH MHC MOLECULE HLA-C*1601 AND USBS
; NUMBER OF INVENTION: THEREOF
; NUMBER OF SEQUENCES: 9

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA

; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch 1.44 Mb storage diskette

; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/038,328
; FILING DATE:

; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/389,360

; FILING DATE:
; APPLICATION DATA:
; APPLICATION NUMBER: 08/079,110

; FILING DATE: June 17, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Pasqualini, Patricia A.

; REGISTRATION NUMBER: 34,894
; REFERENCE/DOCKET NUMBER: LUD 5310.2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200

; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 6:

; SEQUENCE CHARACTERISTICS:
; LENGTH: 18
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear

US-09-038-328-6

Query Match 71.1%; Score 12.8; DB 3; Length 18;
Best Local Similarity 87.5%; Pred. No. 9.6e+02;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 CAGAGAGAGAGGCC 18
DB 18 CAGTGAAGAGAGGCC 3

RESULT 15
US-09-038-328-6

; Sequence 6, Application US/09038328
; Patent No. 6110694
; GENERAL INFORMATION:
; APPLICANT: van der Bruggen et al.
; TITLE OF INVENTION: ISOLATED PEPTIDE WHICH FORMS COMPLEXES
; TITLE OF INVENTION: WITH MHC MOLECULE HLA-C*1601 AND USBS
; NUMBER OF INVENTION: THEREOF
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch 1.44 Mb storage diskette
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/038,328
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/389,360
; FILING DATE:
; APPLICATION DATA:
; APPLICATION NUMBER: 08/079,110
; FILING DATE: June 17, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Pasqualini, Patricia A.
; REGISTRATION NUMBER: 34,894
; REFERENCE/DOCKET NUMBER: LUD 5310.2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 18
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear

OY 3 CAGAGAGAGAGACC 18
|||||
Db 17 CAGAGTGAGACGAGCC 2

Search completed: November 9, 2002, 15:37:14
Job time : 69 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 9, 2002, 14:59:36 ; Search time 84 Seconds

(Without alignments)
76.025 Million cell updates/sec

Title: US-09-902-772-1_COPY_645_662

Perfect score: 18
Sequence: 1 ctcagagagagagagcc 18

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 320260 segs, 177392727 residues

Total number of hits satisfying chosen parameters: 183406

Minimum DB seq length: 0

Maximum DB seq length: 100

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications_NA.*

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match Length	ID	Description
1	14	77.8	19	10	US-09-969-373-3925 Sequence 3925, Ap
2	13.8	76.7	17	10	US-09-866-108-2619 Sequence 2619, Ap
3	13.8	76.7	25	10	US-09-866-108-5548 Sequence 5548, Ap
4	13.8	76.7	25	10	US-09-866-108-5549 Sequence 5549, Ap
5	13.8	76.7	25	10	US-09-866-108-5550 Sequence 5550, Ap
6	13.8	76.7	25	10	US-09-866-108-5551 Sequence 5551, Ap
7	13.8	76.7	25	10	US-09-866-108-5552 Sequence 5552, Ap
8	13.8	76.7	25	10	US-09-866-108-5553 Sequence 5553, Ap
9	13.8	76.7	25	10	US-09-866-108-5554 Sequence 5554, Ap
10	13.8	76.7	25	10	US-09-866-108-5555 Sequence 5555, Ap
11	13.8	76.7	25	10	US-09-866-108-5556 Sequence 5556, Ap
12	13.8	74.4	99	10	US-09-864-761-18548 Sequence 18548, A
13	13	72.2	88	10	US-09-878-574-7158 Sequence 7158, Ap
14	13	72.2	99	10	US-09-969-373-205 Sequence 205, App
15	12.8	71.1	17	10	US-09-866-108-2618 Sequence 2618, Ap
16	12.8	71.1	17	10	US-09-866-108-2620 Sequence 2620, Ap
17	12.8	71.1	17	10	US-09-866-108-6229 Sequence 6229, Ap
18	12.8	71.1	17	10	US-09-866-108-6230 Sequence 6230, Ap
19	12.8	71.1	25	10	US-09-866-108-5547 Sequence 5547, Ap

20	12.8	71.1	25	10	US-09-866-108-5557 Sequence 5557, Ap
21	12.8	71.1	25	10	US-09-866-108-11121 Sequence 11121, A
22	12.8	71.1	25	10	US-09-866-108-11122 Sequence 11122, A
23	12.8	71.1	25	10	US-09-866-108-11123 Sequence 11123, A
24	12.8	71.1	25	10	US-09-866-108-11124 Sequence 11124, A
25	12.8	71.1	25	10	US-09-866-108-11125 Sequence 11125, A
26	12.8	71.1	25	10	US-09-866-108-11126 Sequence 11126, A
27	12.8	71.1	25	10	US-09-866-108-11127 Sequence 11127, A
28	12.8	71.1	25	10	US-09-866-108-11128 Sequence 11128, A
29	12.8	71.1	25	10	US-09-866-108-11129 Sequence 11129, A
30	12.8	71.1	25	10	US-09-866-108-11130 Sequence 11130, A
31	12.4	68.9	20	8	US-08-983-605-91 Sequence 91, Appl
32	12.4	68.9	84	10	US-09-864-761-32975 Sequence 32975, A
33	12.4	68.9	87	12	US-10-078-777-9 Sequence 9, Appl1
34	12.4	68.9	100	10	US-09-969-373-1152 Sequence 1152, Ap
35	12.2	67.8	47	10	US-09-901-484A-225 Sequence 225, App
36	12.2	67.8	51	10	US-09-989-002-53 Sequence 53, Appl
37	12.2	67.8	99	10	US-09-969-373-625 Sequence 625, App
38	12	66.7	83	10	US-09-969-373-742 Sequence 742, App
39	12	66.7	96	10	US-09-969-373-1169 Sequence 1169, Ap
40	12	66.7	96	12	US-10-078-777-8 Sequence 8, Appl1
41	12	66.7	99	10	US-09-969-373-167 Sequence 167, App
42	11.8	65.6	17	10	US-09-866-108-2617 Sequence 2617, Ap
43	11.8	65.6	17	10	US-09-866-108-2621 Sequence 2621, Ap
44	11.8	65.6	17	10	US-09-866-108-6228 Sequence 6228, Ap
45	11.8	65.6	17	10	US-09-866-108-6231 Sequence 6231, Ap

ALIGNMENTS

RESULT 1
US-09-969-373-3925
; Sequence 3925, Application US/09969373
; Patent No. US20020133852A1
; GENERAL INFORMATION:
; APPLICANT: Efferitz, Roger J.
; TITLE OF INVENTION: Soybean SSRs and Methods of Genotyping
; FILE REFERENCE: 38-10(52679)A
; CURRENT APPLICATION NUMBER: US/09/969,373
; PRIOR FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: US 09/754,853
; PRIOR FILING DATE: 2001-01-05
; PRIOR APPLICATION NUMBER: US 09/760,427
; PRIOR FILING DATE: 2001-01-13
; PRIOR APPLICATION NUMBER: US 09/855,768
; PRIOR FILING DATE: 2001-05-15
; NUMBER OF SEQ ID NOS: 4593
; SEQ ID NO 3925
; LENGTH: 19
; TYPE: DNA
; ORGANISM: Glycine max
US-09-969-373-3925

Query Match 77.8%; Score 14; DB 10; Length 19;
Best local Similarity 100.0%; Pred. No. 3; 9e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 AGAGAGAGAGAGC 17
|||
Db 5 AGAGAGAGAGAGC 18

RESULT 2
US-09-866-108-2619
; Sequence 2619, Application US/09866108
; Patent No. US2002048800A1
; GENERAL INFORMATION:
; APPLICANT: GU, Yizhong
; APPLICANT: GU, Yonggang
; APPLICANT: PENN, Sharon G.
; APPLICANT: HANZEL, David K.

```

; APPLICANT: RANK, David R.
; APPLICANT: CHEN, Wensheng
; APPLICANT: SHANNON, Mark
; TITLE OF INVENTION: MYOSIN-LIKE GENE EXPRESSED IN HUMAN HEART AND MUSCLE
; FILE REFERENCE: AEOMICA-7
; CURRENT APPLICATION NUMBER: US/09/866,108
; CURRENT FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 60/266,860
; PRIOR FILING DATE: 2001-02-05
; NUMBER OF SEQ ID NOS: 15752
; SOFTWARE: Aeomica Sequence Listing Engine
; SEQ ID NO 2619
; LENGTH: 17
; TYPE: DNA
; ORGANISM: Homo sapiens
;
US-09-866-108-2619

Query Match      76.7%; Score 13.8; DB 10; Length 17;
Best Local Similarity 88.2%; Pred. No. 4.8e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTCAGAGAGAGAGAGC 17
   |||||  |||||  |||
Db 1 CTCAGAGAGAGAGAGC 17

RESULT 3
US-09-866-108-5548
; Sequence 5548, Application US/09866108
; Patent No. US20020048800A1
; GENERAL INFORMATION:
; APPLICANT: GU, Yizhong
; APPLICANT: JI, Yonggang
; APPLICANT: PENN, Sharon G.
; APPLICANT: HANZEL, David K.
; APPLICANT: RANK, David R.
; APPLICANT: CHEN, Wensheng
; APPLICANT: SHANNON, Mark
; TITLE OF INVENTION: MYOSIN-LIKE GENE EXPRESSED IN HUMAN HEART AND MUSCLE
; FILE REFERENCE: AEOMICA-7
; CURRENT APPLICATION NUMBER: US/09/866,108
; CURRENT FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: GB 24263.6
```

```

; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 60/266,860
; PRIOR FILING DATE: 2001-02-05
; NUMBER OF SEQ ID NOS: 15752
; SOFTWARE: Aeomica Sequence Listing Engine
; SEQ ID NO 5548
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapiens
;
US-09-866-108-5548

Query Match      76.7%; Score 13.8; DB 10; Length 25;
Best Local Similarity 88.2%; Pred. No. 5e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTCAGAGAGAGAGAGC 17
   |||||  |||||  |||
Db 9 CTCAGAGAGAGAGAGC 25

RESULT 4
US-09-866-108-5549
; Sequence 5549, Application US/09866108
; Patent No. US20020048800A1
; GENERAL INFORMATION:
; APPLICANT: GU, Yizhong
; APPLICANT: JI, Yonggang
; APPLICANT: PENN, Sharon G.
; APPLICANT: HANZEL, David K.
; APPLICANT: RANK, David R.
; APPLICANT: CHEN, Wensheng
; APPLICANT: SHANNON, Mark
; TITLE OF INVENTION: MYOSIN-LIKE GENE EXPRESSED IN HUMAN HEART AND MUSCLE
; FILE REFERENCE: AEOMICA-7
; CURRENT APPLICATION NUMBER: US/09/866,108
; CURRENT FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
```

```
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 60/266,860
; PRIOR FILING DATE: 2001-02-05
; NUMBER OF SEQ ID NOS: 15752
; SOFTWARE: Aeomica Sequence Listing Engine
; SEQ ID NO 5549
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-866-108-5549
```

```
Query Match      76.7%; Score 13.8; DB 10; Length 25;
Best Local Similarity 88.2%; Pred. No. 5e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY      1 CTCAGAGAGAGAGAGC 17
          |||||  |||  |||
Db       8 CTCAGAGAGAGAGAGC 24
```

```
RESULT 5
US-09-866-108-5550
; Sequence 5550, Application US/09866108
; Patent No. US20020048800A1
; GENERAL INFORMATION:
; APPLICANT: GU, Yizhong
; APPLICANT: JI, Yonggang
; APPLICANT: PENN, Sharon G.
; APPLICANT: HANZEL, David K.
; APPLICANT: RANK, David R.
; APPLICANT: CHEN, Wensheng
; APPLICANT: SHANNON, Mark
; TITLE OF INVENTION: MYOSIN-LIKE GENE EXPRESSED IN HUMAN HEART AND MUSCLE
; FILE REFERENCE: AEOMICA-7
; CURRENT APPLICATION NUMBER: US/09/866,108
; CURRENT FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
```

```
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 60/266,860
; PRIOR FILING DATE: 2001-02-05
; NUMBER OF SEQ ID NOS: 15752
; SOFTWARE: Aeomica Sequence Listing Engine
; SEQ ID NO 5550
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-866-108-5550
```

```
Query Match      76.7%; Score 13.8; DB 10; Length 25;
Best Local Similarity 88.2%; Pred. No. 5e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY      1 CTCAGAGAGAGAGAGC 17
          |||||  |||  |||
Db       7 CTCAGAGAGAGAGAGC 23
```

```
RESULT 6
US-09-866-108-5551
; Sequence 5551, Application US/09866108
; Patent No. US20020048800A1
; GENERAL INFORMATION:
; APPLICANT: GU, Yizhong
; APPLICANT: JI, Yonggang
; APPLICANT: PENN, Sharon G.
; APPLICANT: HANZEL, David K.
; APPLICANT: RANK, David R.
; APPLICANT: CHEN, Wensheng
; APPLICANT: SHANNON, Mark
; TITLE OF INVENTION: MYOSIN-LIKE GENE EXPRESSED IN HUMAN HEART AND MUSCLE
; FILE REFERENCE: AEOMICA-7
; CURRENT APPLICATION NUMBER: US/09/866,108
; CURRENT FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 60/266,860
; PRIOR FILING DATE: 2001-02-05
; NUMBER OF SEQ ID NOS: 15752
; SOFTWARE: Aeomica Sequence Listing Engine
; SEQ ID NO 5551
```

LENGTH: 25
TYPE: DNA
ORGANISM: Homo sapiens
US-09-866-108-5551

Query Match 76.7%; Score 13.8; DB 10; Length 25;
Best Local Similarity 88.2%; Pred. No. 5e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTCAGAGAGAGAGC 17
||||||| ||| |||
DB 6 CTCAGAGAGAGAGC 22

RESULT 7
US-09-866-108-5552
; Sequence 5552, Application US/09866108
; Patent No. US20020048800A1
; GENERAL INFORMATION:
; APPLICANT: GU, Yizhong
; APPLICANT: JI, Yonggang
; APPLICANT: PENN, Sharron G.
; APPLICANT: HANZEL, David K.
; APPLICANT: RANK, David R.
; APPLICANT: CHEN, Wensheng
; APPLICANT: SHANNON, Mark
; TITLE OF INVENTION: MYOSIN-LIKE GENE EXPRESSED IN HUMAN HEART AND MUSCLE
; FILE REFERENCE: AEWICA-7
; CURRENT APPLICATION NUMBER: US/09/866,108
; CURRENT FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 60/266,860
; PRIOR FILING DATE: 2001-02-05
; NUMBER OF SEQ ID NOS: 15752
; SOFTWARE: Aewmica Sequence Listing Engine
; SEQ ID NO 5552
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-866-108-5552

Query Match 76.7%; Score 13.8; DB 10; Length 25;
Best Local Similarity 88.2%; Pred. No. 5e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTCAGAGAGAGAGC 17

DB 5 CTCAGAGAGAGAGC 21
||||||| ||| |||

RESULT 8
US-09-866-108-5553
; Sequence 5553, Application US/09866108
; Patent No. US20020048800A1
; GENERAL INFORMATION:
; APPLICANT: GU, Yizhong
; APPLICANT: JI, Yonggang
; APPLICANT: PENN, Sharron G.
; APPLICANT: HANZEL, David K.
; APPLICANT: RANK, David R.
; APPLICANT: CHEN, Wensheng
; APPLICANT: SHANNON, Mark
; TITLE OF INVENTION: MYOSIN-LIKE GENE EXPRESSED IN HUMAN HEART AND MUSCLE
; FILE REFERENCE: AEWICA-7
; CURRENT APPLICATION NUMBER: US/09/866,108
; CURRENT FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 60/266,860
; PRIOR FILING DATE: 2001-02-05
; NUMBER OF SEQ ID NOS: 15752
; SOFTWARE: Aewmica Sequence Listing Engine
; SEQ ID NO 5553
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-866-108-5553

Query Match 76.7%; Score 13.8; DB 10; Length 25;
Best Local Similarity 88.2%; Pred. No. 5e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTCAGAGAGAGAGC 17
||||||| ||| |||
DB 4 CTCAGAGAGAGAGC 20

RESULT 9
US-09-866-108-5554
; Sequence 5554, Application US/09866108
; Patent No. US20020048800A1
; GENERAL INFORMATION:
; APPLICANT: GU, Yizhong

```

; APPLICANT: JI, Yonggang
; APPLICANT: PENN, Sharon G.
; APPLICANT: HANZEL, David K.
; APPLICANT: RANK, David R.
; APPLICANT: CHEN, Wensheng
; APPLICANT: SHANNON, Mark
; TITLE OF INVENTION: MYOSIN-LIKE GENE EXPRESSED IN HUMAN HEART AND MUSCLE
; FILE REFERENCE: AEOMICA-7
; CURRENT APPLICATION NUMBER: US/09/866,108
; CURRENT FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 60/266,860
; PRIOR FILING DATE: 2001-02-05
; NUMBER OF SEQ ID NOS: 15752
; SOFTWARE: Aeomica Sequence Listing Engine
; SEQ ID NO 5554
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-866-108-5554

Query Match      76.7%; Score 13.8; DB 10; Length 25;
Best Local Similarity 88.2%; Pred. No. 5e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY      1 CTCAGAGAGAGAGAGC 17
Db      3 CTCAGAGAGCGAGCAGC 19

RESULT 10
US-09-866-108-5555
; Sequence 5555, Application US/09866108
; Patent No. US20020048800A1
; GENERAL INFORMATION:
; APPLICANT: GU, Yizhong
; APPLICANT: JI, Yonggang
; APPLICANT: PENN, Sharon G.
; APPLICANT: HANZEL, David K.
; APPLICANT: RANK, David R.
; APPLICANT: CHEN, Wensheng
; APPLICANT: SHANNON, Mark
; TITLE OF INVENTION: MYOSIN-LIKE GENE EXPRESSED IN HUMAN HEART AND MUSCLE
; FILE REFERENCE: AEOMICA-7
; CURRENT APPLICATION NUMBER: US/09/866,108
; CURRENT FILING DATE: 2001-05-25
```

```

; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 60/266,860
; PRIOR FILING DATE: 2001-02-05
; NUMBER OF SEQ ID NOS: 15752
; SOFTWARE: Aeomica Sequence Listing Engine
; SEQ ID NO 5555
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-866-108-5555

Query Match      76.7%; Score 13.8; DB 10; Length 25;
Best Local Similarity 88.2%; Pred. No. 5e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY      1 CTCAGAGAGAGAGAGC 17
Db      2 CTCAGAGAGCGAGCAGC 18

RESULT 11
US-09-866-108-5556
; Sequence 5556, Application US/09866108
; Patent No. US20020048800A1
; GENERAL INFORMATION:
; APPLICANT: GU, Yizhong
; APPLICANT: JI, Yonggang
; APPLICANT: PENN, Sharon G.
; APPLICANT: HANZEL, David K.
; APPLICANT: RANK, David R.
; APPLICANT: CHEN, Wensheng
; APPLICANT: SHANNON, Mark
; TITLE OF INVENTION: MYOSIN-LIKE GENE EXPRESSED IN HUMAN HEART AND MUSCLE
; FILE REFERENCE: AEOMICA-7
; CURRENT APPLICATION NUMBER: US/09/866,108
; CURRENT FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
```

PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 60/266,860
PRIOR FILING DATE: 2001-02-05
NUMBER OF SEQ ID NOS: 15752
SOFTWARE: Aecomica Sequence Listing Engine
SEQ ID NO 5556
LENGTH: 25
TYPE: DNA
ORGANISM: Homo sapiens
US-09-866-108-5556

Query Match 76.7%; Score 13.8; DB 10; Length 25;
Best Local Similarity 88.2%; Pred. No. 5e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTCAGAGAGAGAGC 17
DB 1 CTCAGAGAGAGAGC 17

RESULT 12
US-09-864-761-18548
Sequence 18548, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 18548
LENGTH: 99
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AL049812.3
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.4
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.1
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.5
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.7
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.1
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.9
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2
OTHER INFORMATION: EST_HUMAN HIT: BE009795.1, EVALU8 2.00e-02
OTHER INFORMATION: NT HIT: U80017.1, EVALU8 8.00e-13
US-09-864-761-18548

Query Match 74.4%; Score 13.4; DB 10; Length 99;
Best Local Similarity 93.3%; Pred. No. 8.5e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TCAGAGAGAGAGAG 16
DB 63 TAAAGAGAGAGAGAG 77

-RESULT 13
US-09-878-574-7158
Sequence 7158, Application US/09878574
Patent No. US20020110548A1
GENERAL INFORMATION:
APPLICANT: Byrum, Joseph R.
APPLICANT: La Rosa, Thomas J.
APPLICANT: Thompson, Michael D.
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
FILE REFERENCE: 38-21(15401)B
CURRENT APPLICATION NUMBER: US/09/878,574
CURRENT FILING DATE: 2001-12-21
PRIOR APPLICATION NUMBER: 09/333,535
PRIOR FILING DATE: 1999-06-14
NUMBER OF SEQ ID NOS: 15775
SEQ ID NO 7158
LENGTH: 88
TYPE: DNA
ORGANISM: Glycine max
OTHER INFORMATION: Clone ID: 701099516H1
US-09-878-574-7158

Query Match 72.2%; Score 13; DB 10; Length 88;
Best Local Similarity 100.0%; Pred. No. 1.3e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 CAGAGAGAGAGGA 15

Db 13 CAGAGAGAGAGA 25

RESULT 14

US-09-969-373-205/C
; Sequence 205, Application US/09969373
; Patent No. US20020133852A1
; GENERAL INFORMATION:
; APPLICANT: Effertz, Roger J.
; APPLICANT: Hauge, Brian M.
; TITLE OF INVENTION: Soybean SSRs and Methods of Genotyping
; FILE REFERENCE: 38-10(52679)A
; CURRENT APPLICATION NUMBER: US/09/969,373
; PRIOR FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: US 09/754,853
; PRIOR FILING DATE: 2001-01-05
; PRIOR APPLICATION NUMBER: US 09/760,427
; PRIOR FILING DATE: 2001-01-13
; PRIOR APPLICATION NUMBER: US 09/855,768
; PRIOR FILING DATE: 2001-05-15
; NUMBER OF SEQ ID NOS: 4593
; SEQ ID NO 205
; LENGTH: 99
; TYPE: DNA
; ORGANISM: Glycine max
US-09-969-373-205

Query Match 72.2%; Score 13; DB 10; Length 99;
Best Local Similarity 100.0%; Pred. No. 1.3e+03;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCAGAGAGAGAG 13
Db 68 CTCAGAGAGAGAG 56

RESULT 15

US-09-866-108-2618
; Sequence 2618, Application US/09866108
; Patent No. US20020048800A1
; GENERAL INFORMATION:
; APPLICANT: GU, Yizhong
; APPLICANT: JI, Yonggang
; APPLICANT: PENN, Sharon G.
; APPLICANT: HANZEL, David K.
; APPLICANT: RANK, David R.
; APPLICANT: CHEN, Wensheng
; APPLICANT: SHANNON, Mark
; TITLE OF INVENTION: MYOSIN-LIKE GENE EXPRESSED IN HUMAN HEART AND MUSCLE
; FILE REFERENCE: AEOMICA-7
; CURRENT APPLICATION NUMBER: US/09/866,108
; CURRENT FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 60/266,860
; PRIOR FILING DATE: 2001-02-05
; NUMBER OF SEQ ID NOS: 15752
; SOFTWARE: Aeomica Sequence Listing Engine
; SEQ ID NO 2618
; LENGTH: 17
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-866-108-2618

Query Match 71.1%; Score 12.8; DB 10; Length 17;
Best Local Similarity 87.5%; Pred. No. 1.4e+03;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTCAGAGAGAGAG 16
Db 2 CTCAGAGAGAGAG 17

Search completed: November 9, 2002, 16:48:17
Job time : 85 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 9, 2002, 14:03:26 ; Search time 2173 Seconds
(without alignments)
134.155 Million cell updates/sec

Title: US-09-902-772-1_COPY_645_662
Perfect score: 18
Sequence: 1 ctcagagagagagagcc 18

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues
Total number of hits satisfying chosen parameters: 357874

Minimum DB seq length: 0
Maximum DB seq length: 100

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estda:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_estl:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pin:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	15.4	85.6	79	AZ647872	AZ647872 IM0514C10
2	15	83.3	95	A1955060	A1955060 wg60c05.x
3	14.6	81.1	84	CNS0349M	AL227155 Tetradon
4	14	77.8	95	AZ952772	AZ952772 2M0217B14
5	13.4	74.4	46	A1697098	A1697098 tq14e10.x
6	13.4	74.4	47	AZ649857	AZ649857 IM0519008

C 7	13.4	74.4	49	9	AA86434	AA86434 o152d09.s
C 8	13.4	74.4	54	17	AZ996539	AZ996539 2M0282F14
C 9	13.4	74.4	75	14	N85423	N85423 J3490F Huma
C 10	13.4	74.4	84	10	BE324070	BE324070 NF013B03P
C 11	13.4	74.4	91	13	BM302259	BM302259 MCA040E10
C 12	13.4	74.4	97	14	B0765595	B0765595 BR003_S0
C 13	13.4	74.4	99	17	AZ325265	AZ325265 1M047006
C 14	13.4	74.4	100	9	AA585224	AA585224 KTH157 HT
C 15	13.2	73.3	23	17	AZ783377	AZ783377 2M0025I04
C 16	13.2	73.3	86	12	BG405995	BG405995 sac40F12.
C 17	13.2	73.3	95	17	AZ489480	AZ489480 1M0321L22
C 18	13.2	73.3	95	17	AL764683	AL764683 Arabidops
C 19	13.2	73.3	100	9	AA470378	AA470378 net10c02.s
C 20	13	72.2	18	12	BG668047	BG668047 DRABU12
C 21	13	72.2	21	17	AZ371089	AZ371089 1M0122D06
C 22	13	72.2	40	9	AA922878	AA922878 o150909.s
C 23	13	72.2	40	13	BI252040	BI252040 602952695
C 24	13	72.2	42	17	AZ590801	AZ590801 1M0400N04
C 25	13	72.2	50	9	AU107978	AU107978 AU107978
C 26	13	72.2	50	9	AU107979	AU107979 AU107979
C 27	13	72.2	52	12	BF651086	BF651086 NF100H09E
C 28	13	72.2	54	13	BM395183	BM395183 50072_-2-7
C 29	13	72.2	56	17	AZ983943	AZ983943 2M0256G18
C 30	13	72.2	57	12	BF640154	BF640154 NF028B011
C 31	13	72.2	60	14	BQ767486	BQ767486 BR008_S0
C 32	13	72.2	68	13	BM521350	BM521350 sal13g09.
C 33	13	72.2	69	17	CNS02XF4	AL219577 Tetradon
C 34	13	72.2	71	9	AA575416	AA575416 v187d09.r
C 35	13	72.2	75	17	AZ506280	AZ506280 1M0347B18
C 36	13	72.2	76	14	BQ094017	BQ094017 f254604.y
C 37	13	72.2	78	10	AM696113	AM696113 NF102F02S
C 38	13	72.2	79	14	R29186	R29186 F1_-295D 22
C 39	13	72.2	79	13	L8315BX	AL354416 Leishman1
C 40	13	72.2	81	13	BM301739	BM301739 MCA044A08
C 41	13	72.2	81	17	AQ073887	AQ073887 EP(3)3288
C 42	13	72.2	84	17	BH223723	BH223723 1006114F0
C 43	13	72.2	85	9	AI152770	AI152770 uc94f01.r
C 44	13	72.2	85	17	BH225000	BH225000 100612380
C 45	13	72.2	90	17	AQ025170	AQ025170 EP(3)1082

ALIGNMENTS

RESULT 1
LOCUS AZ647872 79 bp DNA linear GSS 14-DEC-2000
DEFINITION IM0514C10R Mouse 10kb plasmid UGCI1M library Mus musculus genomic
clone UGCI1M0514C10 R. DNA sequence.
ACCESSION AZ647872
VERSION AZ647872.1 GI:11779772
KEYWORDS GSS.

SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 79)
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, F., Kelly,
M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.
and Wright, D., Weiss, R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)

JOURNAL
COMMENT Contact: Robert B. Weiss
University of Utah
Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA

Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert length: 10000 Std Error: 0.00

Plate: 0514 row: C column: 10
Seq primer: CACACAGGAAACAGCTATACCC
Class: Plasmid ends
High quality sequence stop: 79.
Location/Qualifiers

Trace considered overall poor quality
Seq primer: 40UP from Gibco
High quality sequence stop: 1.
Location/Qualifiers

1. 79
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="U06C1M0514C10"
/clone_1lb="Mouse 10kb plasmid U06C1M 1library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PMD42my; Purified genomic DNA from M. musculus C57BL/6J (male); was obtained from the Jackson Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PMD42 (q14/3211419b1AF129072.1), a copy-number inducible derivative of Plasmid RI. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT
ORIGIN
21 a 14 c 24 g 20 t

Query Match
Best Local Similarity 94.1%; Score 15.4; DB 17; Length 79;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TCAGAGAGAGAGAGCC 18
||||| |||||||
Db 15 TCAGAGAGAGAGAGCC 31

RESULT 2
LOCUS A1955060 95 bp mRNA linear EST 19-AUG-1999
DEFINITION wq60c05.x1 NCI CGAP GC6 Homo sapiens cDNA clone IMAGE:2475656 3', similar to gb:X76981 ADENOSINE A3 RECEPTOR (HUMAN);, mRNA sequence.
ACCESSION A1955060
VERSION A1955060.1 GI:5747382
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 95)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
Unpublished (1997)
Contact: Robert Strussberg, Ph.D.
Email: cga@bcr-remail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/ILMI at: www-bio.lnl.gov/btrp/image/image.html

83.3%; Score 15; DB 9; Length 95;
Best Local Similarity 100.0%; Pred. No. 8.5e+03;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCAGAGAGAGAGA 15
||||| |||||||
Db 29 CTCAGAGAGAGAGA 43

RESULT 3
LOCUS CNS0349M 84 bp DNA linear GSS 15-MAY-2000
DEFINITION Tetradon nigroviridis genome survey sequence PUC-ori end of clone 211H19 of library G from Tetradon nigroviridis, genomic survey sequence.
ACCESSION AL227155
VERSION AL227155.1 GI:7886101
KEYWORDS GSS; genome survey sequence.
SOURCE Tetradon nigroviridis.
ORGANISM Tetradon nigroviridis.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetradon.
1 (bases 1 to 84)
Roest-Crollius, R., Jallou, O., Dasilva, C., Bouneau, L., Fisher, C., Bernot, A., Fitzames, C., Winkler, P., Brothier, P., Quetier, F., Saurin, W. and Weissenbach, J.
Human gene number estimate provided by genome wide analysis using Tetradon nigroviridis DNA sequence
Unpublished
2 (bases 1 to 84)
Roest-Crollius, R., Jallou, O., Dasilva, C., Fitzames, C., Fisher, C., Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and Weissenbach, J.
Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetradon nigroviridis
Unpublished
3 (bases 1 to 84)
Genoscope.
Direct Submission
Submitted (12-APR-2000)
This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetradon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/tetradon.
Location/Qualifiers
1. 84
/organism="Tetradon nigroviridis"
/db_xref="taxon:99883"

FEATURES
source

/clone="211b19"
/note="Genoscope sequence ID : COAG211CD10SP1-end :
PDC-0r1"

BASE COUNT 17 a 23 c 34 g 9 t 1 others

ORIGIN

Query Match 81.1%; Score 14.6; DB 17; Length 84;
Best Local Similarity 88.2%; Pred. No. 1.2e+04;
Matches 15; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 TCAGAGAGAGAGAGACC 18
|||||:|||||
Db 45 TCAGABAGGAGAGAGACC 61

RESULT 4
A2952772/c 95 bp DNA linear GSS 27-APR-2001

LOCUS
DEFINITION 2M0217B14R Mouse 10kb plasmid UGCG2M library Mus musculus genomic
clone UGCG2M0217B14 R. DNA sequence.

ACCESSION
A2952772
VERSION
A2952772.1 GI:13823999
KEYWORDS
GSS.
SOURCE
house mouse.
ORGANISM
Mus musculus.
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 95)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,
M., Rose,M., Rose,R., Stokes,R., Tinney,A., von Niederhausern,A.
and Wright,D.,Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10Kb
Plasmid Inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: dunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0217 row: B column: 14
Seq primer: CACACAGAAACAGCATGACCC
Class: plasmid ends
High quality sequence stop: 95.
Location/Qualifiers

FEATURES
source
1. 95
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UGCG2M0217B14"
/clone_1lb="Mouse 10kb plasmid UGCG2M library"
/sex="Female"
/lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PMD42nv; Purified genomic DNA from M.
musculus C57BL/6J (female) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PMD42 (g14732114gb/AP129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and

purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

BASE COUNT 5 a 40 c 7 g 43 t

ORIGIN

Query Match 77.8%; Score 14; DB 17; Length 95;
Best Local Similarity 100.0%; Pred. No. 2.3e+04;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 CAGAGAGAGAGAGAG 16
|||||:|||||
Db 27 CAGAGAGAGAGAGAG 14

RESULT 5
A1697098 46 bp mRNA linear EST 14-DEC-1999

LOCUS
DEFINITION t914e10.x1 NCI-CGAP-Ut3 Homo sapiens cDNA clone IMAGE:2208810 3,
similar to TR:007611 007611 PROLINE-RICH PROTEOGLYCAN PRPG2.
; contains TAR1.b2 MSRI repetitive element ;, mRNA sequence.

ACCESSION
A1697098
VERSION
A1697098.1 GI:4984998
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 46)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/ILNI at:
www-bio.lnl.gov/bdrp/image/image.html

Trace considered overall poor quality
Insert Length: 1168 Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 1.
Location/Qualifiers

FEATURES
source
1. 46
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2208810"
/clone_1lb="NCI-CGAP-Ut3"
/tissue_type="poorly-differentiated endometrial
adenocarcinoma, 2 pooled tumors"
/lab_host="DH10B"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site:1; Salt;
Site:2; Notti; Cloned unidirectionally. Primer: Oligo df.
Average insert size 1.45 kb. Life Technologies catalog #:
11541-018"

BASE COUNT 10 a 18 c 15 g 3 t

ORIGIN

Query Match 74.4%; Score 13.4; DB 9; Length 46;
Best Local Similarity 93.3%; Pred. No. 2.9e+04;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 4 AGAGAGAGAGAGAGACC 18
|||||:|||||
Db 11 AGAGAGAGAGAGAGACC 25

RESULT 6
 AZ649857/c
 LOCUS
 DEFINITION 47 bp DNA linear GSS 14-DEC-2000
 accession
 AZ649857
 clone UGCG1M0519008 R, DNA sequence.
 accession
 AZ649857.1 GI:11783758
 GSS.
 SOURCE
 house mouse.
 ORGANISM
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 47)
 Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C.,
 Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,
 M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.
 and Wright, D., Weiss, R.
 Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts
 Unpublished (2000)
 Contact: Robert B. Weiss
 University of Utah Genome Center
 University of Utah
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0519 row: 0 column: 08
 Seq primer: CACACAGGAACAGCTATGCACC
 Class: plasmid ends
 High quality sequence stop: 47.
 Location/Qualifiers
 1..47
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UGCG1M0519008"
 /clone_1lb="Mouse 10kb plasmid UGCG1M library"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: pMD42nv; Purified genomic DNA from M.
 musculus C57BL/6J (male) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The
 adaptor DNA was purified and size-selected for a 9.5 to
 10.5 kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of pMD42 (9114732114|9b|AF129072.1), a copy number
 inducible derivative of plasmid R1. The vector was ligated
 with adaptors complementary to the insert adaptors and
 purified. The sheared, adaptor mouse DNA was annealed to
 adaptor vector DNA, and transformed into
 chemically-competent E. coli XL10-Gold (Stratagene) cells
 and selected for ampicillin resistance."

BASE COUNT 2 a 22 c 4 g 19 t
 ORIGIN
 Query Match 74.4%; Score 13.4; DB 17; Length 47;
 Best Local Similarity 93.3%; Pred. No. 2.9e+04;
 Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 4 ACAGAGAGAGAGGCC 18
 ||||||||||||
 DB 27 ACAGAGAGAGATCC 13

RESULT 7
 AA886434/c
 LOCUS
 DEFINITION 49 bp mRNA linear EST 30-MAR-1998
 accession
 o52d09.s1 NCI-CCAP_K1d3 Homo Sapiens cDNA clone IMAGE:1501937 3'
 similar to TR:Q04118 Q04118 SALIVARY PROTEIN-RICH GLYCOPROTEIN G1
 PRECURSOR; contains MER22.t3 MER22 repetitive element ; mRNA
 sequence.
 accession
 AA886434
 clone UGCG2M0282P14 R, DNA sequence.
 accession
 AA886434.1 GI:3001542
 EST.
 SOURCE
 human.
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 1 (bases 1 to 49)
 NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cga@b-riemail.nih.gov
 Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CCAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/ULNL at:
 www.bio.lnl.gov/db/ri/image/image.html

Trace considered overall poor quality
 Seq primer: -40m13 fwd. Ef from Amersham
 High quality sequence stop: 1.
 Location/Qualifiers
 1..49
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1501937"
 /clone_1lb="NCI-CCAP_K1d3"
 /lab_host="DH10B"
 /note="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with
 a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
 strand cDNA was primed with a Not I - oligo(dT) primer,
 double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not
 I and Eco RI sites of the modified pT73 vector. mRNA
 source: 2 pooled kidneys. Library went through one round
 of normalization. Library constructed by Bento Soares and
 M. Fatima Bonaldo."

BASE COUNT 2 a 19 c 21 g 7 t
 ORIGIN
 Query Match 74.4%; Score 13.4; DB 9; Length 49;
 Best Local Similarity 93.3%; Pred. No. 3e+04;
 Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 4 ACAGAGAGAGAGGCC 18
 |||||
 DB 29 ACAGAGAGAGAGGCC 15

RESULT 8
 AZ996539/c
 LOCUS
 DEFINITION 54 bp DNA linear GSS 27-APR-2001
 accession
 2M0282P14R Mouse 10kb plasmid UGCG2M library Mus musculus genomic
 clone UGCG2M0282P14 R, DNA sequence.
 accession
 AZ996539
 clone UGCG2M0282P14 R, DNA sequence.
 accession
 AZ996539.1 GI:13867766
 GSS.
 SOURCE
 house mouse.
 ORGANISM
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 54)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Rellly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SIC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert length: 10000 Std Error: 0.00
Plate: 0282 row: F column: 14
Seq primer: CACACAGGAACAGCATATGACC
Class: plasmid ends
High quality sequence stop: 54.
Location/Qualifiers
1. 54
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUCG2M0282F14"
/clone_1lb="Mouse 10kb plasmid UUCG2M library"
/sex="Female"
/lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (female) was obtained from the Jackson Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g114732114|g114732114|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 1 a 21 c 5 g 27 t
ORIGIN

Query Match 74.4%; Score 13.4; DB 17; Length 54;
Best Local Similarity 93.3%; Pred. No. 3.1e+04;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 4 AGAGAGAGAGAGGCC 18
Db 17 AGAGAGAGAGAGAGTC 3

RESULT 9
N85423 75 bp mRNA linear EST 01-APR-1996
LOCUS N85423
DEFINITION J3490F Human fetal heart, Lambda ZAP Express Homo sapiens cDNA
ACCESSION N85423
VERSION N85423.1 GI:1261048
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 75)
AUTHORS Liew,C.C.
TITLE cDNAs from fetal heart (1996)
JOURNAL Unpublished (1996)
COMMENT Contact: Liew CC
Brigham and Women's Hospital
Harvard Medical School
75 Francis St. Boston, MA 02115, USA
Tel: 6177328915
Fax: 6179750995
Email: cliw@rics.bwh.harvard.edu
Seq primer: GAAATTAACCTCATTAAGGG.
Location/Qualifiers
1. 75
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="J3490"
/clone_1lb="Human fetal heart, Lambda ZAP Express"
/lab_host="E. coli XL1-Blue"
/note="Vector: Lambda ZAP Express; Site_1: EcoRI; Site_2: XhoI; mRNA was purified from human fetal hearts (8-10 weeks). cDNA was synthesized using a XhoI-Oligo dT adaptor-primer. EcoRI adaptors were ligated, followed by digestion with XhoI, for directional cloning into predigested lambda ZAP Express."

BASE COUNT 33 a 4 c 36 g 2 t
ORIGIN

Query Match 74.4%; Score 13.4; DB 14; Length 75;
Best Local Similarity 93.3%; Pred. No. 3.7e+04;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 4 AGAGAGAGAGAGGCC 18
Db 33 AGAGAGAGAGAGGAC 47

RESULT 10
BE324070/c 84 bp mRNA linear EST 21-DEC-2000
LOCUS BE324070
DEFINITION NF013B03PL11025 Phosphate starved leaf Medicago truncatula cDNA
ACCESSION BE324070
VERSION NF013B03PL 5', mRNA sequence.
KEYWORDS BE324070.2 GI:11967233
SOURCE EST.
ORGANISM barrel medic.
Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae; Medicago.

REFERENCE 1 (bases 1 to 84)
AUTHORS Liu,J., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J., Flores,H.R., Inman,J.T., Weller,J.W., May,G.D. and Harrison,M.J.
TITLE Expressed Sequence Tags from the Samuel Roberts Noble Foundation Medicago truncatula phosphate-starved leaf library
JOURNAL Unpublished (2000)
COMMENT On Jul 14, 2000 this sequence version replaced gi:19197847.
Contact: Harrison MJ
Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73402, USA
Tel: 580 221 7325
Fax: 580 221 7380
Email: mjharrison@noble.org
Medicago Genome Initiative accession: MGI:S:21524
Insert length: 884 Std Error: 0.00
Plate: 013 row: B column: 03
Seq primer: TCACACAGGAACAGCATATGAC.
Location/Qualifiers
1. 84
/organism="Medicago truncatula"

BASE COUNT	20 a	29 c	11 g	24 t
ORIGIN	/db.xref="taxon:3880" /clone="NF013803PL" /clone_lib="Phosphate starved leaf" /tissue_type="leaf" /dev_stage="trifoliolate" /note="Vector: Lambda Zap; At the trifoliolate stage, M. truncatula plants were transplanted to phosphate-free sand and grown for a further 30 days. During this 30 day period, the plants were fertilized twice weekly with 1/2 Hoaglands solution containing only 20uM potassium phosphate. RNA was prepared from above ground tissues."			
Query Match	74.4%;	Score 13.4;	DB 10;	Length 84;
Best Local Similarity	93.3%;	Pred. No. 3.9e+04;		
Matches 14;	Conservative	0;	Mismatches 1;	Indels 0;
Gaps 0;				
OY	2	TCAGAGAGAGAGAG 16		
Db	48	TTAGAGAGAGAGAG 34		
RESULT 11				
BM302259/c				
LOCUS	BM302259	91 bp	mRNA	linear
DEFINITION	MCA040E10.23702 Ice plant Lambda Uni-Zap XR expression library, 0			EST 22-JAN-2002
	hours NaCl treatment prescreened for removal of highly abundant			
	transcripts Mesembryanthemum crystallinum cDNA clone MCA040E10 5,			
	mRNA sequence.			
ACCESSION	BM302259			
VERSION	BM302259.1	GI:18024634		
KEYWORDS	EST.			
SOURCE	Common Iceplant.			
ORGANISM	Mesembryanthemum crystallinum			
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;			
	Caryophyllidae; Caryophyllales; Alzooceae; Mesembryanthemum.			
REFERENCE	1 (bases 1 to 91)			
AUTHORS	Cushman, J.C.			
TITLE	An expressed sequence tag database for the common ice plant,			
	Mesembryanthemum crystallinum			
	Unpublished (1997)			
JOURNAL	Contact: Cushman JC			
COMMENT	Department of Biochemistry			
	University of Nevada			
	MS200, Reno, NV 89557-0014, USA			
	Tel: 775-784-1918			
	Fax: 775-784-1650			
	Email: jcushman@unr.edu			
	PCR Primers			
	FORWARD: T3 20mer			
	BACKWARD: T7 21mer			
	Plate: 040 row: E column: 10			
	Seq primer: T3 20mer			
	High quality sequence stop: 91.			
FEATURES	Location/Qualifiers			
source	1..91			
	/organism="Mesembryanthemum crystallinum"			
	/db_xref="taxon:3544"			
	/clone="MCA040E10"			
	/clone_lib="Ice plant Lambda Uni-Zap XR expression library			
	/clone_lib="0 hours NaCl treatment prescreened for removal of highly			
	abundant transcripts"			
	/tissue_type="leaf"			
	/dev_stage="six-week-old"			
	/note="Vector: Lambda Uni-Zap XR, Bluescript SK-; Site_1:			
	ECORI; Site_2: XhoI"			
BASE COUNT	16 a	12 g	36 t	
ORIGIN				
Query Match	74.4%;	Score 13.4;	DB 13;	Length 91;
Best Local Similarity	93.3%;	Pred. No. 4e+04;		

Matches	14;	Conservative	0;	Mismatches	1;	Indels	0;	Gaps	0;
OY	3	CAGAGAGAGAGAGACC	17						
LOCUS									
DEFINITION									
ACCESSION									
VERSION									
KEYWORDS									
ORGANISM									
SOURCE									
REFERENCE									
AUTHORS									
TITLE									
JOURNAL									
COMMENT									
FEATURES									
SOURCE									
Matches	14;	Conservative	0;	Mismatches	1;	Indels	0;	Gaps	0;
OY	3	CAGAGAGAGAGAGACC	17						
LOCUS									
DEFINITION									
ACCESSION									
VERSION									
KEYWORDS									
ORGANISM									
SOURCE									
REFERENCE									
AUTHORS									
TITLE									
JOURNAL									
COMMENT									
FEATURES									
SOURCE									
Matches	14;	Conservative	0;	Mismatches	1;	Indels	0;	Gaps	0;
OY	3	CAGAGAGAGAGAGACC	17						
LOCUS									
DEFINITION									
ACCESSION									
VERSION									
KEYWORDS									
ORGANISM									
SOURCE									
REFERENCE									
AUTHORS									
TITLE									
JOURNAL									
COMMENT									
FEATURES									
SOURCE									
Matches	14;	Conservative	0;	Mismatches	1;	Indels	0;	Gaps	0;
OY	3	CAGAGAGAGAGAGACC	17						
LOCUS									
DEFINITION									
ACCESSION									
VERSION									
KEYWORDS									
ORGANISM									
SOURCE									
REFERENCE									
AUTHORS									
TITLE									
JOURNAL									
COMMENT									
FEATURES									
SOURCE									
Matches	14;	Conservative	0;	Mismatches	1;	Indels	0;	Gaps	0;
OY	3	CAGAGAGAGAGAGACC	17						
LOCUS									
DEFINITION									
ACCESSION									
VERSION									
KEYWORDS									
ORGANISM									
SOURCE									
REFERENCE									
AUTHORS									

AUTHORS Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Relilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A. and Wright, D., Weis, R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

JOURNAL Unpublished (2000)

COMMENT Contact: Robert B. Weiss
University of Utah
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0047 row: 0 column: 06
Seq primer: CACACAGGAACAGCATATGACC
Class: plasmid ends
High quality sequence stop: 99.

FEATURES
source
1. .99
/location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="U06C1M0047006"
/clone_lib="Mouse 10kb plasmid U06C1M library"
/sex="Male"
/lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PMD29ny. Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g1147321149b1AF129072.1), a copy-number inducible derivative of plasmid RL. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 19 a 27 c 25 g 28 t

ORIGIN

Query Match 74.4%; Score 13.4; DB 17; Length 99;
Best Local Similarity 93.3%; Pred. No. 4.2e+04;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTCAGAGAGAGAGA 15
||||| |||||||
Db 67 CTCAGTGAGAGAGA 53

RESULT 14 AA585224 100 bp mRNA linear EST 09-SEP-1997
AA585224/c
LOCUS KTH157 HTCDL1 Homo sapiens cDNA 5'/3' similar to Myosin heavy chain
DEFINITION gene, mRNA sequence.
ACCESSION AA585224
VERSION AA585224.1 GI:2385112
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 100)

AUTHORS Sohn, U., Park, D.S., Lee, C.M., Cho, W.K., Ahn, H.J., Lee, M.Y., Hwang, M.Y. and Jin, S.W.

TITLE Human HTCDL1 library cDNAs

JOURNAL Unpublished (1994)

COMMENT Contact: Uik Sohn, D.S., Park, C.M., Lee, W.K., Cho, H.J., Ahn, M.Y., Lee, M.Y., Hwang, S.W., Jin
Laboratory of Molecular Biology
Kyungpook National University
Dept. of Genetic Eng., Kyungpook National Univ., Taegu 702-701, Korea
Tel: 82-053-950-5382
Fax: 82-053-955-5327
Email: usohn@bn.kyungpook.ac.kr
Seq primer: M13 Reverse/SK primer.

FEATURES
source
1. 100
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="HTCDL1"
/lab_host="XL1-Blue"
/note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2: EcoRI; Poly(A)-mRNA from the 2-year old male fetal thymus, oligo(dT) priming, EcoRI cloning in the vector pBluescript (Stratagene)."

BASE COUNT 23 a 26 c 18 g 33 t

ORIGIN

Query Match 74.4%; Score 13.4; DB 9; Length 100;
Best Local Similarity 93.3%; Pred. No. 4.2e+04;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TCAGAGAGAGAGAG 16
||||| |||||
Db 90 TCAGAGAGAGAGAG 76

RESULT 15 A2783377 23 bp DNA linear GSS 16-FEB-2001
A2783377/c
LOCUS 2M0025104F Mouse 10kb plasmid U06C1M library Mus musculus genomic
DEFINITION clone U06C2M0025104 F, DNA sequence.
ACCESSION A2783377
VERSION A2783377.1 GI:12918045
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 23)

AUTHORS Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Relilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A. and Wright, D., Weis, R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

JOURNAL Unpublished (2000)

COMMENT Contact: Robert B. Weiss
University of Utah
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0025 row: 1 column: 04
Seq primer: CGTTGTAACGACGCCACACT
Class: Plasmid ends
High quality sequence stop: 23.

FEATURES
source
1. 23
/organism="Mus musculus"
/strain="C57BL/6J"

/db_xref="taxon:10090"
 /clone="UUGC2M0025104"
 /clone_lib="Mouse 10kb plasmid UUGC1M library"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: PMD42uv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (<http://www.jax.org/resources/documents/dnares/>). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PMD42 (G14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 1 a 7 c 5 g 10 t
 ORIGIN

Query Match 73.3%; Score 13.2; DB 17; Length 23;
 Best Local Similarity 83.3%; Pred. No. 2.5e+04;
 Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CTCAGAGAGAGAGAGACC 18
 ||||| ||||| ||| |
 Db 19 CTCAGGAGAGAGAGAGAC 2

Search completed: November 9, 2002, 15:35:57
 Job time : 2178 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 9, 2002, 09:10:11 ; Search time 3941.96 Seconds

(without alignments)
10682.939 Million cell updates/sec

Title: US-09-902-772-1

Perfect score: 1447
Sequence: 1 gaattccgcgaacgaataa.....gaaagaaagcgccaagaaaa 1447

Scoring table:

IDENTITY_NTC
Gapop 10.0 , Gapept 1.0

Searched: 2054640 segs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: gb_da:*
2: gb_htg:*
3: gb_in:*
4: gb_cm:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_cm:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_in:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_man:*
37: em_htg_vtl:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1447	100.0	1447	6 E31253	E31253 Protein hav
2	1409	97.4	1413	5 AY065661	AY065661 Gallus ga
3	1356	93.7	1528	6 E31254	E31254 Protein hav
4	1344	92.9	1516	5 GGERG	X77159 G.gallus ER
5	936	64.7	1808	10 AB031088	AB031088 Rattus no
6	921	63.6	3166	9 HUMERG2	M17254 Human ery2
7	889.6	61.5	2012	5 XLAJ4126	AJ224126 Xenopus 1
8	881.6	60.9	2131	10 AB073078	AB073078 Mus muscu
9	867.2	59.9	2209	10 AB073080	AB073080 Mus muscu
10	828	57.2	2133	10 AB073079	AB073079 Mus muscu
11	801.8	55.4	1798	5 XLAJ4125	AJ224125 Xenopus 1
12	760.2	52.5	3126	9 HUMERG11	M21535 Human ery p
13	571.2	39.5	1436	5 DRE249590	AJ249590 Danio rer
14	570.2	39.4	1729	10 MMFLI1	X59421 Mouse Flt-1
15	563.8	39.0	2916	6 A36461	A36461 Sequence 2
16	563.8	39.0	2938	6 AR080101	AR080101 Sequence
17	563.8	39.0	2938	9 HSHDFLI	X67001 H.sapiens H
18	562.2	38.9	1359	9 AY029368	AY029368 Homo sapi
19	562.2	38.9	2394	9 BC010115	BC010115 Homo sapi
20	562.2	38.9	2957	9 HUMERGBFLI	M98833 Homo sapien
21	562.2	38.9	2959	9 BC001670	BC001670 Homo sapi
22	561.4	38.8	6849	12 AY029367	AY029367 Synthetic
23	560.6	38.7	1673	9 S45205	S45205 Flt-1-Frien
24	549.4	38.0	4403	5 XLFILG	X66979 X.laevis MR
25	518.6	35.8	1245	5 AF177538	AF177538 Danio rer
26	514	35.5	3490	5 CCNNAFLI	Y14773 Colurnix co
27	502.2	34.7	3545	5 CCELIONCO	Y14774 Colurnix co
28	493.2	34.1	762	9 S72621	S72621 EMS...ery f
29	478.2	33.0	1932	9 HUMFLI1A	M93255 Human Flt-1
30	477	33.0	816	10 S66169	S66169 Ery-3-Immuno
31	387	26.7	14518	9 AP001423	AP001423 Homo sapi
32	387	26.7	100000	9 AP000021	AP000021 Homo sapi
33	387	26.7	100000	9 AP000163	AP000163 Homo sapi
34	387	26.7	340000	9 AP001731	AP001731 Homo sapi
35	373	25.8	588	9 S72865	S72865 EMS...EMS-e
36	358.6	24.8	1431	9 AF327066	AF327066 Homo sapi
37	320.4	22.1	66352	9 HSY17293	Y17293 Homo sapien
38	320.4	22.1	187131	2 AP001535	AP001535 Homo sapi
39	320.4	22.1	217382	9 AP001122	AP001122 Homo sapi
40	214.4	14.8	634	3 SULERG	M61067 Sea urchin
41	203	14.0	491	3 HDI311813	AJ311813 Hediuste d
42	196.6	13.6	2180	3 AY060316	AY060316 Drosophila
43	180	12.4	1403	9 HSDNAFEV3	Y08979 H.sapiens F
44	180	12.4	1752	10 RND91679	U91679 Rattus norv
45	180	12.4	169741	9 AC097468	AC097468 Homo sapi

ALIGNMENTS

RESULT 1
E31253
LOCUS E31253 1447 bp DNA linear PAT 18-JUN-2001
DEFINITION Protein having cell calcifying inhibitory activity and gene
encoding the same.
ACCESSION E31253
VERSION E31253.1 GI:13025685
KEYWORDS UP 1999075871-R/L.
SOURCE unidentifed.
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 1447)
AUTHORS Hiroyasu, I., Yoshinobu, H., Marjito, P., Joel, R. and Helena, E.
TITLE Protein having cell calcifying inhibitory activity and gene
encoding the same

JOURNAL Patent: JP 1999075871-A 1 23-MAR-1999.
 CHUGAI PHARMACEUT CO LTD, HE TRUSTEES OF THE UNIVERSITY OF
 PENNSYLVANIA
 COMMENT OS Unidentified
 PN JP 1999075871-A/1
 PD 23-MAR-1999
 PE 29-MAY-1998 JP 1998166076
 PR 18-JUN-1997 US 08/878177, 20-JUN-1997 US 60/050297 PI
 HIROYASU IWAMOTO, YOSHINOBU HIGUCHI, MARIJO PASHIFIKI, PI JOEL
 ROZENBURG, PI
 HELENA E
 PI C12N15/09,A61K48/00,C07K14/465,C07K16/18,C12Q1/68//A61K38/00,
 PC C12P21/02*,
 PC C12P21/08,C12N15/00,A61K37/02
 CC Strandedness: Single;
 CC Topology: Linear;
 FH Key
 FT source 1..1447 Location/Qualifiers
 1..1447 Location/Qualifiers
 /organism="Unidentified"
 /db_xref="taxon:32644"
 BASE COUNT 440 a 374 c 317 g 316 t
 ORIGIN

Query Match 100.0%; Score 1447; DB 6; Length 1447;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1447; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAATTCGGGAGGAGATATTTATTATTTAGCAATTTATAGCATTAATCTTGATCA 60
 DB 1 GAATTCGGGAGGAGATATTTATTATTTAGCAATTTATAGCATTAATCTTGATCA 60
 QY 61 TTATGCAAGCACTATTAAAGAAAGCATTTATCACTGCTGATGAAGACCACTCTTTTG 120
 DB 61 TTATGCAAGCACTATTAAAGAAAGCATTTATCACTGCTGATGAAGACCACTCTTTTG 120
 QY 121 AGTGTGCTAGCGAGTGGCCCACTTGGAAAGACAGAAATGACAGCCTCTTCCAGTG 180
 DB 121 AGTGTGCTAGCGAGTGGCCCACTTGGAAAGACAGAAATGACAGCCTCTTCCAGTG 180
 QY 181 AATATGGGAAACATCAAGATGAGAGCCCGCTTCCAGAGCACTGTTATCAGC 240
 DB 181 AATATGGGAAACATCAAGATGAGAGCCCGCTTCCAGAGCACTGTTATCAGC 240
 QY 241 CCCCGGCGAGATTACCTTAAGATGAGTTAAACCAACAGTTAATGGTCAAGA 300
 DB 241 CCCCGGCGAGATTACCTTAAGATGAGTTAATGGTCAAGA 300
 QY 301 AATCACTGATGACCTGACGCTGGCCAAAGAGGAAATGGTTAGCAGTTCAGACATG 360
 DB 301 AATCACTGATGACCTGACGCTGGCCAAAGAGGAAATGGTTAGCAGTTCAGACATG 360
 QY 361 TTGGGATGAATGGAAGCTACATGGAAGAGATATTCGCGCTCCAAATATGACAA 420
 DB 361 TTGGGATGAATGGAAGCTACATGGAAGAGATATTCGCGCTCCAAATATGACAA 420
 QY 421 CCAATGAAGAGAGTTATGTGCCAGAGATCTTACGTTATGAGACAGACATGTAC 480
 DB 421 CCAATGAAGAGAGTTATGTGCCAGAGATCTTACGTTATGAGACAGACATGTAC 480
 QY 481 GGCAGTGGCTGAGTGGGAGTGAAGAGATATGCTTCCAGACGTTGATCTGTTGT 540
 DB 481 GGCAGTGGCTGAGTGGGAGTGAAGAGATATGCTTCCAGACGTTGATCTGTTGT 540
 QY 541 TCCAGAACATTTATGGGAAAGAGTTGTGTAATGACCAAAATGACTTCCAGAGACTCA 600
 DB 541 TCCAGAACATTTATGGGAAAGAGTTGTGTAATGACCAAAATGACTTCCAGAGACTCA 600
 QY 601 CCGCGAGCTATAAGCAGATATCTCTGTCAACACTACACTACCTCAGAGAGAGGAG 660
 DB 601 CCGCGAGCTATAAGCAGATATCTCTGTCAACACTACACTACCTCAGAGAGAGGAG 660

QY 661 CCACCTTTTATTTTCCAAATATACATGATTTACCCAGAGCAACGCAAGAAATACACAA 720
 DB 661 CCACCTTTTATTTTCCAAATATACATGATTTACCCAGAGCAACGCAAGAAATACACAA 720
 QY 721 GGCAGATTTACTTATTTGAGCAACCCAGAGATCACCCTGGAGAGATCAGCATCCCA 780
 DB 721 GGCAGATTTACTTATTTGAGCAACCCAGAGATCACCCTGGAGAGATCAGCATCCCA 780
 QY 781 CTCAGTCAAAAGGTATCCCAACATCATCTTCAACAGCTGCCCCAAACAGAGACAGCTG 840
 DB 781 CTCAGTCAAAAGGTATCCCAACATCATCTTCAACAGCTGCCCCAAACAGAGACAGCTG 840
 QY 841 CTCAGTTAGATCTTATTCAGATTTCTTGGACCAACAGCCGCTTTGCAAAATCCAGGA 900
 DB 841 CTCAGTTAGATCTTATTCAGATTTCTTGGACCAACAGCCGCTTTGCAAAATCCAGGA 900
 QY 901 GTGGGAGATACAGCTATGAGCACTGCTCTACTGAGAGCTTCTGTGAGACAGCTCCAACTCCA 960
 DB 901 GTGGGAGATACAGCTATGAGCACTGCTCTACTGAGAGCTTCTGTGAGACAGCTCCAACTCCA 960
 QY 961 ACTGATCACTGAGAGGACCAAAATGGGAGTTCAAGATGACAGACCTGATGAAGTG 1020
 DB 961 ACTGATCACTGAGAGGACCAAAATGGGAGTTCAAGATGACAGACCTGATGAAGTG 1020
 QY 1021 CTCGGCGTTGGGAGAGAGGAAAGCAAACTTAACATGACATGACAAATCAGCCGTG 1080
 DB 1021 CTCGGCGTTGGGAGAGAGGAAAGCAAACTTAACATGACATGACAAATCAGCCGTG 1080
 QY 1081 CACTTGGCTACTACTATGACAAATAATTTATGCTAAAGTTCTAGTGAACGCTATGCTT 1140
 DB 1081 CACTTGGCTACTACTATGACAAATAATTTATGCTAAAGTTCTAGTGAACGCTATGCTT 1140
 QY 1141 ACNAATTTATTTCCAGGAATGCTCAGAGCCCTCCAGCCTCAGCTCCAGAAATCATCA 1200
 DB 1141 ACNAATTTATTTCCAGGAATGCTCAGAGCCCTCCAGCCTCAGCTCCAGAAATCATCA 1200
 QY 1201 TGTACAAATACCCATCAGACCTCCCTACATGAGTTCTACCATGACACCCCCAGAGA 1260
 DB 1201 TGTACAAATACCCATCAGACCTCCCTACATGAGTTCTACCATGACACCCCCAGAGA 1260
 QY 1261 TGAATCTTTAGCTCCCATCCCTGCTTGGCCCTTAACCTATCCACTTTTGTGCTG 1320
 DB 1261 TGAATCTTTAGCTCCCATCCCTGCTTGGCCCTTAACCTATCCACTTTTGTGCTG 1320
 QY 1321 CCCCTAATCATACTGGAATTCACCAACTGAGAGCATCTACCCCAATACCAAGCTCCAG 1380
 DB 1321 CCCCTAATCATACTGGAATTCACCAACTGAGAGCATCTACCCCAATACCAAGCTCCAG 1380
 QY 1381 CTGCTCATATGCTTCCCATCTTGGCACTACTACTAAGTGGGAAAGAAAGAACGCC 1440
 DB 1381 CTGCTCATATGCTTCCCATCTTGGCACTACTACTAAGTGGGAAAGAAAGAACGCC 1440
 QY 1441 AAGAAA 1447
 DB 1441 AAGAAA 1447

RESULT 2
 AY065661 1413 bp mRNA linear VRT 17-DEC-2001
 LOCUS Gallus gallus erg isoform C-1-1 mRNA, complete cds; alternatively
 DEFINITION
 ACCESSION AY065661
 VERSION AY065661.1 GI:17887440
 KEYWORDS
 SOURCE
 ORGANISM
 Gallus gallus.
 Gallus gallus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Archosauromia; Aves; Neognathae; Galliformes; Phasianidae;
 Phasianinae; Gallus.
 1 (bases 1 to 1413)
 Iwamoto,M., Higuchi,Y., Koyama,E., Enomoto-Iwamoto,M., Kurisu,K.,
 AUTHORS

Yeh, H., Abrams, W. R., Rosenbloom, J. and Pacifici, M.
Transcription factor ERG variants and functional diversification of
chondrocytes during limb long bone development
J. Cell Biol. 150 (1), 27-40 (2000)

JOURNAL 20351415
MEDLINE 10893254
PUBMED 2 (bases 1 to 1413)
REFERENCE Iwamoto, M., Higuchi, Y., Enomoto-Iwamoto, M., Kurisu, K., Koyama, E.,
AUTHORS Yeh, H., Rosenbloom, J. and Pacifici, M.
TITLE The role of ERG (ets related gene) in cartilage development
JOURNAL Osteoarthritis. Cartil. 9 Suppl A, S41-S47 (2001)
MEDLINE 21535378
PUBMED 11680687
REFERENCE 3 (bases 1 to 1413)
AUTHORS Iwamoto, M., Higuchi, Y., Yeh, H. and Pacifici, M.
TITLE Direct Submission
JOURNAL Submitted (05-DEC-2001) Oral Anatomy and Developmental Biology,
Osaka University Faculty of Dentistry, 1-8 Yamadaoka, Suita, Osaka
565-0871, Japan

FEATURES
source Location/Qualifiers
1..1413
/organism="Gallus gallus"
/db_xref="taxon:9031"
37..1392
/note="alternatively spliced; transcription factor"
/codon_start=1
/product="erg isoform C-1-1"
/protein_id="FAL4089.1"
/db_xref="GI:17887441"
/translation="MASTIKELSVSEDSIFECAYSGPHLAKTEMTASSEYQOT
SKMSPRVPODWLSOPPARVTIKMECNPNQVNSRSPDCSVAAGKVVSSDNGM
NYSYMEKHPPMNTNERVIVPADPTLSTHROWLMEVAYEGLPVDILIF
ONIDKELEKMTDFORLTPSYNADILLSHYLREKATFIFPTSYPRATORT
TPDPLPYEBOARSAMTSHSPKOSKATOSSTVPTDEOROLDYOLIGPSSSLA
NCSGOIOCMORLELSDSSNCTMGTEGSEPMMDPVARWBRKRPNNY
DKLSRALRTYIDKNITKTKVGRKRYAKTFDHFIAQLQHPPESSMTKIPSDLPYSS
YHAPQKNFVAPHPALPVTSSSFFAANPYMNSPTGIGIYPTRLPAHMPSHLGT
Y"

BASE COUNT 424 a 370 c 311 g 308 t
ORIGIN

Query Match 97.4%; Score 1409; DB 5; Length 1413;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1409; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 TAGCAATTATTAGCATTAATCTTGATCACAATTATGCAAGCACTATTAAAGAAACA 86
DB 1 TAGCAATTATTAGCATTAATCTTGATCACAATTATGCAAGCACTATTAAAGAAACA 60
QY 87 TTATCAGTGTGAGTGAAGACAGCTCTGTTGAGTGTGCTACGAGATGCCCACTT 146
DB 61 TTATCAGTGTGAGTGAAGACAGCTCTGTTGAGTGTGCTACGAGATGCCCACTT 120
QY 147 GCAAAAGCAGAAATGACAGCTCTCTTCAGATGATATGGCAACATCAAAAGATGAGC 206
DB 121 GCAAAAGCAGAAATGACAGCTCTCTTCAGATGATATGGCAACATCAAAAGATGAGC 180
QY 207 CGCGGGTCCCGACAGAGAGCTGTTATCAGACCCCGGCGCAGAGTTACCAATTAAGATG 266
DB 181 CGCGGGTCCCGACAGAGAGCTGTTATCAGACCCCGGCGCAGAGTTACCAATTAAGATG 240
QY 267 GAGGTAAACCAACAGGTTAATGAGTCAAGAAATTCACCTGATGACTGACAGCTGAGCA 326
DB 241 GAGGTAAACCAACAGGTTAATGAGTCAAGAAATTCACCTGATGACTGACAGCTGAGCA 300
QY 327 AAAGAGGAGAAATGTTAGCAGTCAAGCAATGTTGGATGAGCACTATGGAAGCTACATG 386
DB 301 AAAGAGGAGAAATGTTAGCAGTCAAGCAATGTTGGATGAGCACTATGGAAGCTACATG 360
QY 387 GAAGAGAGAGATATTCGGCTCCAAATATGACAAACCAAGTAAGAGAGTTATGTGCCA 446
DB 361 GAAGAGAGAGATATTCGGCTCCAAATATGACAAACCAAGTAAGAGAGTTATGTGCCA 420

QY 447 GCAGATCTTACGTTATGAGACACAGACCATGTACGGCAGTGGCTGAGTGGSCATGAG 506
DB 421 GCAGATCTTACGTTATGAGACACAGACCATGTACGGCAGTGGCTGAGTGGSCATGAG 480
QY 507 GAGTATGCTTTCACAGACCTGGACATCTTGTCTTCAGAACTATGATGGGAAAGATTG 566
DB 481 GAGTATGCTTTCACAGACCTGGACATCTTGTCTTCAGAACTATGATGGGAAAGATTG 540
QY 567 TGTAAATGACCAAGATGACTTCCAGAGACTACGCGGAGCTATATAAGCAGATATCTC 626
DB 541 TGTAAATGACCAAGATGACTTCCAGAGACTACGCGGAGCTATATAAGCAGATATCTC 600
QY 627 CTGTACACCTTACACTACTCAGAGAGAGAGAGCCACTTTATTTTCCAAATATCATCA 686
DB 601 CTGTACACCTTACACTACTCAGAGAGAGAGAGCCACTTTATTTTCCAAATATCATCA 660
QY 687 GTTTACCCAGAACCAACGCAAGATTAACAAGGCGACATTTACTTATGAGCAAGCG 746
DB 661 GTTTACCCAGAACCAACGCAAGATTAACAAGGCGACATTTACTTATGAGCAAGCG 720
QY 747 AGGAGATCAGCGTGGACAGTCAAGCCATCCCACTCAGTCAAAAAGCTTACCAACATCA 806
DB 721 AGGAGATCAGCGTGGACAGTCAAGCCATCCCACTCAGTCAAAAAGCTTACCAACATCA 780
QY 807 TCTTCAACAGTGGCCCAAAACAGAAAGACACGCTCTCAGTTAGATCCTTATCAGATTCTT 866
DB 781 TCTTCAACAGTGGCCCAAAACAGAAAGACACGCTCTCAGTTAGATCCTTATCAGATTCTT 840
QY 867 GGACCGACGACGCGCTCTTGGCAAAATCCAGGAGTGGGAGATATACGCTATGGCAATTTC 926
DB 841 GGACCGACGACGCGCTCTTGGCAAAATCCAGGAGTGGGAGATATACGCTATGGCAATTTC 900
QY 927 CTCTGAGAGTTTGTGTCGACAGCTCCCACTCAACTGCACTACCTGGGAGGCGACAAT 986
DB 901 CTCTGAGAGTTTGTGTCGACAGCTCCCACTCAACTGCACTACCTGGGAGGCGACAAT 960
QY 987 GGGGAGTTCAAGATGACAGACCCCTGATGAATGGCTCGCGTTGGGAGAGAGAAAAAGC 1046
DB 961 GGGGAGTTCAAGATGACAGACCCCTGATGAATGGCTCGCGTTGGGAGAGAGAAAAAGC 1020
QY 1047 AAACCTTACATGAACTTATGACAAACCTACAGCGCTGCTACTGCTATGACAAATAT 1106
DB 1021 AAACCTTACATGAACTTATGACAAACCTACAGCGCTGCTACTGCTATGACAAATAT 1080
QY 1107 ATTATGACTAAAGTTCTATGTTAAACGCTATGCTTACAAATTTGATTTCCACGGAATGCT 1166
DB 1081 ATTATGACTAAAGTTCTATGTTAAACGCTATGCTTACAAATTTGATTTCCACGGAATGCT 1140
QY 1167 CAGGCGCTCCAGCTCACCTCCAGAAATCATTCATGACAAATACCATCAGACCTCCCG 1226
DB 1141 CAGGCGCTCCAGCTCACCTCCAGAAATCATTCATGACAAATACCATCAGACCTCCCG 1200
QY 1227 TACATGATTCCTTACCATGACACACCCCAAGATGAACTTTGATAGTCCCATCCCTT 1286
DB 1201 TACATGATTCCTTACCATGACACACCCCAAGATGAACTTTGATAGTCCCATCCCTT 1260
QY 1287 GCTTTGCGCGTAAACCTTATGAGTAAACGCTTATGCTGCGCCCTTATCCATCTGGAATTTACCA 1346
DB 1261 GCTTTGCGCGTAAACCTTATGAGTAAACGCTTATGCTGCGCCCTTATCCATCTGGAATTTACCA 1320
QY 1347 ACTGAGAGCATCTACCCCAATACAGGCTGCCAGCTGCTCATATGCTTCCCATCTTGGC 1406
DB 1321 ACTGAGAGCATCTACCCCAATACAGGCTGCCAGCTGCTCATATGCTTCCCATCTTGGC 1380
QY 1407 ACCTACTACTTAAGTGGGGAAGAAAGAAA 1435
DB 1381 ACCTACTACTTAAGTGGGGAAGAAAGAAA 1409

RESULT 3
E31254 1528 bp DNA linear PAT 18-JUN-2001
LOCUS E31254 Protein having cell calcifying inhibitory activity and gene
DEFINITION

encoding the same.

ACCESSION E31254
 VERSION E31254.1 GI:13025686
 KEYWORDS JP 199075871-A/2.
 SOURCE unidentified.
 ORGANISM unidentified.
 REFERENCE 1 (bases 1 to 1528)
 AUTHORS Hiroyasu, I., Yoshinobu, H., Marijo, P., Joel, R. and Helena, E.
 TITLE Protein having cell calcifying inhibitory activity and gene
 encoding the same
 JOURNAL Patent: JP 199075871-A 2 23-MAR-1999;
 CHUGAI PHARMACEUT CO LTD, HE TRUSTEES OF THE UNIVERSITY OF
 PENNSYLVANIA
 COMMENT OS Unidentified
 PN JP 199075871-A/2
 PD 23-MAR-1999
 PE 29-MAY-1998 JP 1998166076
 PR 18-JUN-1997 US 08/878177, 20-JUN-1997 US 60/050297 PI
 HIROYASU IWAMOTO, YOSHINOBU HIGUCHI, MARIJO PASHITKI, PI JOEL
 ROZENBURG, PI
 PI HELENA E
 PC C12N15/09, A61K48/00, C07K14/465, C07K16/18, C1201/68//A61K38/00,
 PC C12P21/02,
 PC C12P21/08, C12N15/00, A61K37/02
 CC Strandedness: Single;
 CC Topology: Linear;
 FH key
 FT Location/Qualifiers
 FT source 1.1528
 FEATURES Location/Qualifiers
 source 1.1528
 /organism="unidentified"
 /db_xref="taxon:32644"
 BASE COUNT 464 a 395 c 330 g 339 t
 ORIGIN
 Query Match 93.7%; Score 1356; DB 6; Length 1528;
 Best Local Similarity 94.7%; Pred. No. 0;
 Matches 1447; Conservative 0; Mismatches 0; Indels 81; Gaps 1;

481 GGCAGTGGCTGGAGTGGGCGATGGAAGAGTATGGTCTTCCACAGCTGGACATCTGTGTCT 540
 |||||||
 Db 481 GGCAGTGGCTGGAGTGGGCGATGGAAGAGTATGGTCTTCCACAGCTGGACATCTGTGTCT 540
 |||||||
 QY 541 TCCAGAACATTTGATGGGAAGAGTTGTGTAATGACCAAGATGACTTCCAGAGACTCA 600
 |||||||
 Db 541 TCCAGAACATTTGATGGGAAGAGTTGTGTAATGACCAAGATGACTTCCAGAGACTCA 600
 |||||||
 QY 601 GCCCGAGCTATTAACGCGAGATATCTCTGTCAACCTACACTACCTCAGAGAGA----- 654
 |||||||
 Db 601 GCCCGAGCTATTAACGCGAGATATCTCTGTCAACCTACACTACCTCAGAGAGCTCTC 660
 |||||||
 QY 655 ----- 654
 |||||||
 Db 661 TTCACATTTGACTTCAGATGATGTTGATTAAGGCGCTTACAAAACCTCCACGGTTAATGC 720
 |||||||
 QY 655 -----GAGGAGGCACCTTTATTTTCCAAATCATCAGTTTCCAGAG 699
 |||||||
 Db 721 ATGCTAAGAAACACAGAGAGAGCCACTTTTATTTTCCAAATCATCAGTTTACCCAGAG 780
 |||||||
 QY 700 CACGCAAAAGATTAACAACAAAGCCAGATTTACTTTATGAGCAGAGAGAGATCAGCCT 759
 |||||||
 Db 781 CACGCAAAAGATTAACAACAAAGCCAGATTTACTTTATGAGCAGAGAGAGATCAGCCT 840
 |||||||
 QY 760 GGACGAGTACAGCCATCCCTACTCAGTCAAAAGCTAACCAACATCATCTTAAACAGTGC 819
 |||||||
 Db 841 GGACGAGTACAGCCATCCCTACTCAGTCAAAAGCTAACCAACATCATCTTAAACAGTGC 900
 |||||||
 QY 820 CCAAAAACAAAGACAGCGCTCTCAGTTTATGATCTTATGATGATTTCTTGAGACGACAGCA 879
 |||||||
 Db 901 CCAAAAACAAAGACAGCGCTCTCAGTTTATGATCTTATGATGATTTCTTGAGACGACAGCA 960
 |||||||
 QY 880 GCCGCTTGGCAATCCAGGAGTGGGCGAGATACAGCTATGAGCAGTTCCTACTGAGCTTC 939
 |||||||
 Db 961 GCCGCTTGGCAATCCAGGAGTGGGCGAGATACAGCTATGAGCAGTTCCTACTGAGCTTC 1020
 |||||||
 QY 940 TGTGGAGACGCTCCCAATCTCAATCTGATCAGCTGGGAGGCGCAAAATGGGGAGTTCAAGA 999
 |||||||
 Db 1021 TGTGGAGACGCTCCCAATCTCAATCTGATCAGCTGGGAGGCGCAAAATGGGGAGTTCAAGA 1080
 |||||||
 QY 1000 TGCACAGCCCTGATGAAGTGGCTGCGCGTTGGGAGAGAGAGAAAGCAAACTTAACATGA 1059
 |||||||
 Db 1081 TGCACAGCCCTGATGAAGTGGCTGCGCGTTGGGAGAGAGAGAAAGCAAACTTAACATGA 1140
 |||||||
 QY 1060 ACTATGACAAACTCAGCCGTGCACTTCCTACTACTATGACAAAAATTTATGACTAAAG 1119
 |||||||
 Db 1141 ACTATGACAAACTCAGCCGTGCACTTCCTACTACTATGACAAAAATTTATGACTAAAG 1200
 |||||||
 QY 1120 TTCTATGSTRAAAGCTATGCTTCAAAATTTGATTTCCAGGGAATGCTCAGGCCCTCCAGC 1179
 |||||||
 Db 1201 TTCTATGSTRAAAGCTATGCTTCAAAATTTGATTTCCAGGGAATGCTCAGGCCCTCCAGC 1260
 |||||||
 QY 1180 CTCACCTTCGAATATCATCATGCTACAAATATCCATCAGAGACTCCCTTACATGAGTTCT 1239
 |||||||
 Db 1261 CTCACCTTCGAATATCATCATGCTACAAATATCCATCAGAGACTCCCTTACATGAGTTCT 1320
 |||||||
 QY 1240 ACCATGCAACCCCGAGAGATGAATGAACTTTGAGCTCCCATCCCTCTCTTTCGCCGTAA 1299
 |||||||
 Db 1321 ACCATGCAACCCCGAGAGATGAATGAACTTTGAGCTCCCATCCCTCTCTTTCGCCGTAA 1380
 |||||||
 QY 1300 CCTCATCAAGCTTTTCTGCTGCCCTAATCCATACCTGGAATTCACCAACTGGAGGCACT 1359
 |||||||
 Db 1381 CCTCATCAAGCTTTTCTGCTGCCCTAATCCATACCTGGAATTCACCAACTGGAGGCACT 1440
 |||||||
 QY 1360 ACCCAATACAGAGCTGCGAGCTGCTCATATGCCCTTCCATCTTGGCAGCTACTCTAAG 1419
 |||||||
 Db 1441 ACCCAATACAGAGCTGCGAGCTGCTCATATGCCCTTCCATCTTGGCAGCTACTCTAAG 1500
 |||||||
 QY 1420 TGGGAGAAAGAAAGAGCGCCAAAGAAA 1447
 |||||||
 Db 1501 TGGGAGAAAGAAAGAGCGCCAAAGAAA 1528
 |||||||

RESULT 4

GENE GGERG 1516 bp mRNA linear VRT 27-APR-1995

LOCUS G.gallus ERG mRNA. x77159

DEFINITION G.gallus ERG mRNA.

ACCESSION x77159

VERSION x77159.1 GI:790439

KEYWORDS ERG gene.

SOURCE chicken.

ORGANISM Gallus gallus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianidae; Gallus.

REFERENCE 1 (bases 1 to 1516)

AUTHORS Dhordain, P., Dewitte, F., Desbiens, X., Stehelin, D. and Dutertre, C.

TITLE Mesodermal expression of the chicken erg gene associated with precartilaginous condensation and cartilage differentiation

JOURNAL Mech. Dev. 50 (1), 17-28 (1995)

MEDLINE 95329425

PUBMED 7605748

REFERENCE 2 (bases 1 to 1516)

AUTHORS Dutertre, C.

TITLE Direct Submission

JOURNAL Submitted (17-JAN-1994) M. Dutertre-Cocquilland, CNRS UA 1160, Oncologie Molculaire, Institut Pasteur, 1 rue Calmette, 59019 Lille, FRANCE

FEATURES

source location/Qualifiers

1. 1516

/organism="Gallus gallus"

/db_xref="taxon:9031"

/issue_type="spleen"

/clone_lib="lambda gt10"

/dev_stage="adult"

63. 1499

/gene="ERG"

63. 1499

/gene="ERG"

/codon_start=1

/protein_id="CA54404.1"

/db_xref="GI:790440"

/db_xref="SWISS-PROT:Q90837"

/translation="MASTIKELSVSEEDSLFECAVSGSPHLAKTEMTASSSSEVGT SKSPVROODMLSOPARVITIKMECPNOVNSRSPDSCSAKKGKSSSDNNGV NGSYVEKHIPPNNITNERIVTADPTLWSTDEVRKMLEVAVEYGLPDVILF ONIDGELCKMDDPRLTPSYNADILSLHLYRETPPLHITSDVVKALONSRL MAHNTGATFEPNTPSYPEATORTTPDLPBQARSAMTSHSPQSKATOPSS SYVKTEDORPOLDPYOLIGPTSSRLANGSGIOLMOTLELSDSSNSCITWGT NGEFKMTDDEVARBMGERSKPNMYDRLRYVYDKNIMTYHGRKRYAKPPEH GIAQALQHPHPRESMKRTPSDLPYMSYIAHPOKNFVAHPHAPLVTSSEFAADNP YMNSPTGGITPNTRLPAAHPHSLGTY"

BASE COUNT 458 a 392 c 327 g 339 t

ORIGIN

Query Match 92.9%; Score 1344; DB 5; Length 1516;

Best Local Similarity 94.7%; Pred. No. 0;

Matches 143; Conservative 0; Mismatches 0; Indels 81; Gaps 1;

QY 241 CCCCCGAGAGTTACATTAAGATGATGATTAACCCAAACAGGTTAATGGCTCAAGCA 300

DB 241 CCCCCGAGAGTTACATTAAGATGATGATTAACCCAAACAGGTTAATGGCTCAAGCA 300

QY 301 ATTACCGTGAATGAGAGGCTGGCAAAAGAGGGAATGGTTAGAGATTCACACATG 360

DB 301 ATTACCGTGAATGAGAGGCTGGCAAAAGAGGGAATGGTTAGAGATTCACACATG 360

QY 361 TTGGGATGAATGAGAGGCTGGCAAAAGAGGGAATGGTTAGAGATTCACACATG 420

DB 361 TTGGGATGAATGAGAGGCTGGCAAAAGAGGGAATGGTTAGAGATTCACACATG 420

QY 421 CCAATGACGAGAGGTTATGTCGACAGATCTCTACGTTATGAGCAGACCATGTAC 480

DB 421 CCAATGACGAGAGGTTATGTCGACAGATCTCTACGTTATGAGCAGACCATGTAC 480

QY 481 GGCAGTGGCTGAGAGGCTGGCAAAAGAGGGAATGGTTAGAGATTCACACATG 540

DB 481 GGCAGTGGCTGAGAGGCTGGCAAAAGAGGGAATGGTTAGAGATTCACACATG 540

QY 541 TCCAGACATTTGATGAGAGAGGTTGTTAAATGACCAAAAGATGACTTCCAGAGACTCA 600

DB 541 TCCAGACATTTGATGAGAGAGGTTGTTAAATGACCAAAAGATGACTTCCAGAGACTCA 600

QY 601 CGCCGAGCTATTAACGAGATATCTCTCTGACACCTACCTACCTCAGAGAGA----- 654

DB 601 CGCCGAGCTATTAACGAGATATCTCTCTGACACCTACCTACCTCAGAGAGAGCTCTC 660

QY 655 ----- 654

DB 655 ----- 654

QY 661 TTCACATTTGACTTCAGATGATGATTAAGGCTTACAAACTCTCACAGGTTAATGC 720

DB 661 TTCACATTTGACTTCAGATGATGATTAAGGCTTACAAACTCTCACAGGTTAATGC 720

QY 655 -----GAGAGGACCTTTATTTTTCAAATACATCAGTTTACCCAGAG 659

DB 655 -----GAGAGGACCTTTATTTTTCAAATACATCAGTTTACCCAGAG 659

QY 721 ATGCTAGAAACAGAGAGGAGGACCTTTATTTTTCAAATACATCAGTTTACCCAGAG 780

DB 721 ATGCTAGAAACAGAGAGGAGGACCTTTATTTTTCAAATACATCAGTTTACCCAGAG 780

QY 700 CAACGCAAGATTAACACAAAGGCGAGATTTACCTTATGAGCAAGCAGAGAGATCAGCGT 759

DB 700 CAACGCAAGATTAACACAAAGGCGAGATTTACCTTATGAGCAAGCAGAGAGATCAGCGT 759

QY 781 CAACGCAAGATTAACACAAAGGCGAGATTTACCTTATGAGCAAGCAGAGAGATCAGCGT 840

DB 781 CAACGCAAGATTAACACAAAGGCGAGATTTACCTTATGAGCAAGCAGAGAGATCAGCGT 840

QY 760 GAGAGAGTCAAGGATCCACATCCAGTCAAGCAAGGATACCCACCATCTTCAACATGTC 819

DB 760 GAGAGAGTCAAGGATCCACATCCAGTCAAGCAAGGATACCCACCATCTTCAACATGTC 819

QY 841 GAGAGAGTCAAGGATCCACATCCAGTCAAGCAAGGATACCCACCATCTTCAACATGTC 900

DB 841 GAGAGAGTCAAGGATCCACATCCAGTCAAGCAAGGATACCCACCATCTTCAACATGTC 900

QY 820 CCAAAACAGAGAGCAGAGCTCTCAGTTAGATCTTATGATCTTCTTGAGCCGACGCA 879

DB 820 CCAAAACAGAGAGCAGAGCTCTCAGTTAGATCTTATGATCTTCTTGAGCCGACGCA 879

QY 901 CCAAAACAGAGAGCAGAGCTCTCAGTTAGATCTTATGATCTTCTTGAGCCGACGCA 960

DB 901 CCAAAACAGAGAGCAGAGCTCTCAGTTAGATCTTATGATCTTCTTGAGCCGACGCA 960

QY 880 GCGGCTTGGCAAAATCCAGGAGTGGGCAAGTACAGCTATGCGAGTTCCTTGAAGCTTC 939

DB 880 GCGGCTTGGCAAAATCCAGGAGTGGGCAAGTACAGCTATGCGAGTTCCTTGAAGCTTC 939

QY 961 GCGGCTTGGCAAAATCCAGGAGTGGGCAAGTACAGCTATGCGAGTTCCTTGAAGCTTC 1020

DB 961 GCGGCTTGGCAAAATCCAGGAGTGGGCAAGTACAGCTATGCGAGTTCCTTGAAGCTTC 1020

QY 940 TGTGCGACAGCTCCCACTCCCACTGATCCTACCTGAGGAGGCAAAATGGGAGTTCAAGA 999

DB 940 TGTGCGACAGCTCCCACTCCCACTGATCCTACCTGAGGAGGCAAAATGGGAGTTCAAGA 999

QY 1021 TGTGCGACAGCTCCCACTCCCACTGATCCTACCTGAGGAGGCAAAATGGGAGTTCAAGA 1080

DB 1021 TGTGCGACAGCTCCCACTCCCACTGATCCTACCTGAGGAGGCAAAATGGGAGTTCAAGA 1080

QY 1000 TGAACAGACCTGATGAAAGTGGCTGCGGCTGGGAGAGAGGAAAGCAAACTTAACATGA 1059

DB 1000 TGAACAGACCTGATGAAAGTGGCTGCGGCTGGGAGAGAGGAAAGCAAACTTAACATGA 1059

QY 1081 TGACAGACCTGATGAAAGTGGCTGCGGCTGGGAGAGAGGAAAGCAAACTTAACATGA 1140

DB 1081 TGACAGACCTGATGAAAGTGGCTGCGGCTGGGAGAGAGGAAAGCAAACTTAACATGA 1140

QY 1060 ACTATGACAAAGTCAAGCGGTGCACTTGTCTACTACTATGACAAAAATATTATGACTAAG 1119

DB 1060 ACTATGACAAAGTCAAGCGGTGCACTTGTCTACTACTATGACAAAAATATTATGACTAAG 1119

QY 1141 ACTATGACAAAGTCAAGCGGTGCACTTGTCTACTACTATGACAAAAATATTATGACTAAG 1200

DB 1141 ACTATGACAAAGTCAAGCGGTGCACTTGTCTACTACTATGACAAAAATATTATGACTAAG 1200

QY 1120 TTTATGTTAAAGCTTATGCTTAAATTTATTTTCAAGCAATTCGCTCAGGCTTCAGC 1179

DB 1120 TTTATGTTAAAGCTTATGCTTAAATTTATTTTCAAGCAATTCGCTCAGGCTTCAGC 1179

QY 1201 TTTATGTTAAAGCTTATGCTTAAATTTATTTTCAAGCAATTCGCTCAGGCTTCAGC 1260

DB 1201 TTTATGTTAAAGCTTATGCTTAAATTTATTTTCAAGCAATTCGCTCAGGCTTCAGC 1260

QY 1180 CTCACCTTCAGAGATATCATCATTAACCAATACAGACTCCCTTACATGATGTTCT 1239

DB 1180 CTCACCTTCAGAGATATCATCATTAACCAATACAGACTCCCTTACATGATGTTCT 1239

QY 1261 CTCACCTTCAGAGATATCATCATTAACCAATACAGACTCCCTTACATGATGTTCT 1320

DB 1261 CTCACCTTCAGAGATATCATCATTAACCAATACAGACTCCCTTACATGATGTTCT 1320

QY 1240 ACCATGACACACCCCAAGAGATGAATTTGTAGCTCCCATCCCTGCTTTCGCCGCTAA 1299

QY	1300	CCTCATCCAGCCTTTTGGTGGCCCTATTCATCTGGAATTCACCACTGGAGCATCT	1359
Db	1380	ACCATGACACACCCCGAGAAATGAACTTGTGTAGCTCCCATCCCTCGCTTGGCCGTTAA	1380
QY	1360	ACCCCAATACGAGCGTGCACAGCTGCTCATATAGCCTTCCCATCTTGGACCTTACTAAG	1419
Db	1441	ACCCCAATACGAGCGTGCACAGCTGCTCATATGCTTCCCATCTTGGACCTTACTAAG	1500
QY	1420	TGGGGAAGAAAGAAA 1435	
Db	1501	TGGGGAAGAAAGAAA 1516	
RESULT 5	AB031088		
LOCUS	AB031088	1808 bp	mRNA linear ROD 07-AUG-2001
DEFINITION	Rattus norvegicus VESPI4 mRNA for vascular endothelial cell specific protein 14, complete cds.		
ACCESSION	AB031088		
VERSION	AB031088.1	GI:15128488	
KEYWORDS	vascular endothelial cell specific protein 14.		
SOURCE	Rattus norvegicus liver cDNA to mRNA.		
ORGANISM	Rattus norvegicus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
AUTHORS	1 (bases 1 to 1808) Aoki,T., Toyoda,H., Nishimoto,S., Tawara,J., Ukai,Y. and Komurasaki,T.		
TITLE	Identification of VESPI4,a vascular endothelial cell specific protein		
JOURNAL	Unpublished		
REFERENCES	2 (bases 1 to 1808) Aoki,T., Toyoda,H., Nishimoto,S., Tawara,J., Ukai,Y. and Komurasaki,T.		
AUTHORS	Direct Submission		
JOURNAL	Submitted (12-AUG-1999) Tomomi Aoki, Taisho Pharmaceutical Co., Ltd., Medicinal Research Laboratories, Molecular Biology Lab., 1-403, Yoshino-cho, Omiya, Saitama 3308530, Japan (E-mail:s17704ecm.taisho.co.jp, Tel:048-663-1111, Fax:048-652-7254)		
FEATURES			
SOURCE	location/Qualifiers		
gene	1..1808	location/Qualifiers	
CDS	/organism="Rattus norvegicus" /db_xref="taxon:10116" /tissue_type="liver" 1..1808 /gene="VESPI4" 63..1430 /note="VESPI4" /note="similar to human erg1" /codon_start=1 /product="vascular endothelial cell specific protein 14" /protein_id="BAB62744.1" /db_xref="GI:15128488" /translation="MASTIKFALSVSKDSLEFCATGPHLATENTASSSDYQGT SKSPRYQDQWLSQPPARVITKMECPDSQVNGSRNSPDECSYRGKMGSPVVG SYGSEMEKHMPPNMNTTNERVITVPADPTLMSYDHYKQEMAVKEGLADVLL QNIDGELCKTKDQFQRLTPSYNADILLSHLYLREPLPLHSDVDKALONSPL MHAANDLPEYPPRPSRTWTGSHPTPOSKAQPSPSYVTKEDORPOLDPQIOLG SRLANPGSGIOIRYMDQFLLELSDSSNGCTMETGNETEFKTDDEYVARMGESK NMANDKSRALRYLYDKNIMTKYHGKRYAKFPDHGTAQALQHPPESSLYKTS YMGSYTHHPQAMNFAHPHPLPVTVSSSFATPNPWNPSPTGGLYPMTRILPASH Lgtyr"		
BASE COUNT	510 a 522 c 426 g 350 t		
ORIGIN			
Query Match	64.7%: Score 936; DB 10; Length 1808;		
Best Local Similarity	80.2%: Pred. No. 5.8e-273;		
Matches 1140; Conservative	0; Mismatches 270; Indels 12; Gaps 3;		

QY	8	GCGAACAAATAATTAATTATTAAGCAATTAATTAGCGATCAATTAATCTTGATCAATTAATGAGC	67
Db	8	GAGAGCAATATATTAATTATTAATTAACATTTATTAACGATCAATTAACCTTGATGATTCGATTAATGCG	67
QY	68	AAGCATTAATTAAAGAGCAATTAATCAAGTGGTGAAGCAACAGTCCCTGTTGTTGAAGTGTGC	127
Db	68	CAGCACTAATTAAAGAGGCGCTTGTCAAGTGTGAGCAAGACAGTCACTATTTTGAAGTGTGC	127
QY	128	CTAGAGATGCGCCCACTTCCCAAAAGACAGAAATGACAGGCTCCTCTCTCCAGGAAATATG	187
Db	128	CTAGAGAAAGCCACACTGCGCTAGAGACAGAAATGACCGCATCCTCTCTCCAGGAACTATGG	187
QY	188	GCAAACTCAAAAGATGAGCCCGCGCGCTCCCGACAGAGACTGGTTATCAACACCCCGCGC	247
Db	188	CCAGATACATCCAAAGATGAGTCCAGAGTCCCAACAGAGACTGGTGTCTCAACCCCGACG	247
QY	248	CAGAGTTACCATTTAAGATGGAATGTCTAACCCAAACAGGTTAATGGGTCAAGGAATTCAC	307
Db	248	CAGGGTCACCATCAAGATGGAATGTGCAACCTTACCCAGGTGAACGGTTCAGGAATTCAC	307
QY	308	TGATGATGACGCGTGGCAAAAGAGGAAATAGTTACTAGTCAAGTCAAGATGTTGGGAT	367
Db	308	TGAGAGATGCAAGTGTGACCAAAAGTGGGAAGATGGTGGGACACCTCGATATCTGTGGAAAT	367
QY	368	GAACTATGAGAGCTACATGGAAGAGAACATATTTCCGCTCCAATATGACAAACCAATGA	427
Db	368	GAGCTAGGAGAGCTACATATGGAAGAGAACATGCGCGCCCAATATATGACCAAGAAATGA	427
QY	428	ACGAAAGATTAATGTGGCACAGATCCTCGATTATGAGACACAGACATGTCAGGCACTG	487
Db	428	ACGAGAGTGTATGTCCTCTGCAAGATCTCTATCTGTGACACAGACATGTCGCGACGTG	487
QY	488	GCTGAGATGGGCAAGTGAAGAGATATGATGCTTCACAGACGTGACATCTTGTTGTTCCAAA	547
Db	488	GCTGAGATGGGCAAGTGAAGAAATACGGCCCTCTAGACGTGAGAGTCTTATTAATTTCAGAA	547
QY	548	CATTGATGGGAAAGATGTGTGTAATATGACCAAAAGATGACTTCCAGAGACTACGCCGAG	607
Db	548	TATTGAGGGGAAGAGGTGTGCAAGATGACAAAAGTGAATCTCCAGAGGCTTCTACTCCGAG	607
QY	608	CTATAAGCAGATATCCTCTGTACACACCTACACTACACAGAGAGA-----GAGGAGCC	662
Db	608	CTAAATAGCCGCAATCTTCTCTCAATCTCCACTACAGAGAGACTCCCTTCACAA	667
QY	663	ACTTTATTTTCCAAATATCATCAGTTTACCCAGAA-----GCAACGCAAAAGATTAACAC	718
Db	668	TCTGACATCCGAGAGAGCTGTGATTAAGGCTTTACAAAATCTCCACAGGTTAATGATCATG	727
QY	719	AAGGCGAGATTTACCTTATATGAGCAAGCGAGAGATATAGGTTGAGAGATCAACAGCATCC	778
Db	728	AATACACAGATTTACCTTATATAGGCTCCCGAGAGATTAACCTTGAGCCGCGCACAGGCACTCC	787
QY	779	CA---CTCAGTCAAAAGCTACCCCAACATCATCTTCAACAGTGGCCCAAAACAGAGAGCA	835
Db	788	CACCCCTCAGATCCAAAGCTGCCAGGCAATCTCCCTCCACAGTGGCCCAAACTGAAAGCA	847
QY	836	GCGTCTCAATAGATCCTTATATCAGATTTCTTGAACCGACACAGACGCGTCTTGCAATATCC	895
Db	848	GCGTCTCAATAGATTCCTTATACCAAGATCTCTGGAGCCACCACTATCTGCGCTTGCAATATCC	907
QY	896	AGGAGATGGGCAATACAGCTATGAGAGTTCCTACTGAGAGCTTCTGTGGAGAGCTGCCAA	955
Db	908	AGGTAGTGGGCAATTCAGTGTGTGAGAGTTCTCTGCTAGAACTCTGTGTGACAGCTCCAA	967
QY	956	CTCCAACTGATCACTGTGGAGGCGCAAAATGGGAGTTTCAAGATGACAGACCTGATGA	1015
Db	968	CTCCAACTGATCACTCTGGGGAAGGCGCAACCGGGGAGTTTCAAGATGAGAGACCCGAGTGA	1027
QY	1016	AGTGGCTCGGCGTGTGGGGAGAGAGGAAAGCAAACTTAATGATCACTATGACAAACTAG	1075
Db	1028	GGTGGCTCGGCGCTGGGGGAGGCGGGAAGCAAGCCCAACATGAATCTATGACAAACTAG	1087

RESULT 6	HUMERG2	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
1076	CCGTGACACTTCGCTTACCTACTACTATGACAAAAAATTTTATGACTAAAGTTTCATGTATTAACGCTA	1135	Human <i>erg2</i> gene encoding <i>erg2</i> protein, complete cds.	M17254	GI:182186	Human cell line COLO 320, cDNA to mRNA, clone lambda 12.		
1088	CGGTGACCTCCGCTCTACTACTAGCAAAAAATCTATGACCAAGGTGACAGGGAACGCTA	1147						
1136	TGCTTACAAATTTGATTTTCCAGCAATGGCTGAGGCGCTCCGAGCCCTCCAGCCCTCCAGCAATC	1195						
1148	TGCTTACAAAGTTTGACTTTCACGGGATTTGGCCAGGCGCTCCGAGCCCTCCAGCCCTCCAGCAATC	1207						
1196	ATCCATGTACAAATTTACCATCAGACCTCCCTTACATGATTTCTACCAATGACACGCCCA	1255						
1208	GTCCCTGTACAAAGTACACCTCCGAGCTCCGCTGACATGATGGCTCTTATCACCACGCCCA	1267						
1236	GAGATGAACTTTGAGTCTCCCATCCCTCCGCTTTGGCCCGTAACTCATCCAGCTTTT	1315						
1268	GAAAGTGAAGACTTTGTGGTCCCGACCCCTCCGAGCCCTCCAGTACATCTTCAGATTCTT	1327						
1316	TGCTTACAAATTTGATTTTCCAGCAATGGCTGAGGCGCTCCGAGCCCTCCAGCAATC	1375						
1328	TGCTTACAAAGTACACCTCCGAGCTCCGCTGACATGATGGCTCTTATCACCACGCCCA	1387						
1376	GCCAGCTGCTCATATGATGCTTCCCATCTTGGGACCTACTACTA	1417						
1388	CCGACGACCATATGCGCTCTCAGCTTGGGACCTACTACTA	1429						
REFERENCE								
AUTHORS								
TITLE								
JOURNAL								
MEDLINE								
PUBMED								
FEATURES								
source								
gene								
CDS								
repeat_region								
repeat_region								
polyA_signal								
BASE COUNT	915 a	770 c	768 g	713 t				

Query Match	63.6%	Score 921:	DB 9:	Length 3166:
Best Local Similarity	81.1%	Prod. No. 2.2e-268:		
Matches 1100:	Conservative 0:	Mismatches 245:	Indels 12:	Gaps
ORIGIN	227 bp upstream of PVU1 site; chromosome 21.			
QY	74 TATTAAAGGAACATTTATTCAGTGTGTAGTGAAGACCAAGTCCCTGTTTGTAGTGTGCTTACAGG	133		
DB	289 TATCAAGGAACCTTTATTCAGTGTGTAGTGAAGACCAAGTGTGTTTGTAGTGTGCTTACAGG	348		
QY	134 ATCGCCCCACTCTTCAAAAGACAGAAATGACAGCCTCTCTTCCATGTAAATATGGCAAC	193		
DB	349 AACGCACACCTGTCTAAAGACAGAGATAGCGGCTCTCTCCAGCAGCTATGTAGACAGAC	408		
QY	194 ATCAAAAGTATGAGCCCGGCGCTTCCCGACAGACCTGTTATACACCTCCCGGCGCAAGT	253		
DB	409 TTCCAAAGTATGAGCCCGGCGCTTCCCGACAGATTTGGCTGTCTCAACCCCGCAGCGAGT	468		
QY	254 TACCATTAAGATGAGTGTAAACCCAAACAGTTAAATGGGTCAAGGAATTCACCTGTATGA	313		
DB	469 CACCATCAAAATGTAAATGTAAACCTTACCGCAGTGTAAATGGGTCAAGGAATTCACCTGTATGA	528		
QY	314 CTGCAAGCGTGGCAAAAGAGAGGAAATATGTTAGCACTTCAAGACATATGTTGGATGACTA	373		
DB	529 ATGGAAGTGTGGCCAAAGCGGGAAGATGTGGGACGCCAGACACCGCTTGGATGTAACTA	588		
QY	374 TGGAAAGTACATGTGAAGGAAAGCAATATTCGCGCTCCAAATATATCAACAATGAAGCAAG	433		
DB	589 CGGCAAGCTACATGTGAAGGAAAGCAATATTCGCGCTCCAAATATATCAACAAGCAAGCGAG	648		
QY	434 AGTTATTTGTGGCAGCAGATCTTACGTTATGTAGACACAGACCATGTACGGCAGTGGCTGA	493		
DB	649 AGTTATGTGTGGCAGCAGATCTTACGTTATGTAGATACAGACCATGTACGGCAGTGGCTGA	708		
QY	494 GTGGGCACTGAAGAGATATGTCTTCCACAGCTGGACATCTTGTGTTCCAGACATTTGA	553		
DB	709 GTGGGCGGTGAAGAAATATGTGCGCTTCCAGCTGCAACATCTTGTATTCAGAACATGTGA	768		
QY	554 TGGGAAAGAGTGTGTAAATATGAACAAAGATGACTCCAGAGACATCAAGCCGAGTATTA	613		
DB	769 TGGGAAAGAACTGTGCAAGATGTACCAAGAGCGACTTCCAGAGGCTTACCCCGACGTATCAA	828		
QY	614 CGCAGATATCTCTCTGCACACCTTACACTTACCTCAGAGAGA-----GAGGAGCCAC	664		
DB	829 CGCGCAATCTCTCTCTGCACATCTCCACTACAGAGACTCTCTTCCACATTTGAC	888		
QY	665 TTTTATTTTCCAAATATCATCATGTTTACCAGAGCAAGCAAGAAATATCAACAAGGCC	724		
DB	889 TTTCAAGATGATGTGATTAAGCTTTTACAAAACCTTCCACGGTTAAATGACATGTAGAAACAC	948		
QY	725 AGATTTCCTTATGAGCAACGAGAGATTCAGGTTGAGAGATTCACAGCATCCCA---C	781		
DB	949 AGATTTCCTTATGAGCAACGAGAGATTCAGGTTGAGAGATTCACAGCATCCCA---C	1008		
QY	782 TCACTCAAAAGCTATCCCAACCATCATCTTCAACAGTGGCCCAAAACAGAAAGACACAGCTGC	841		
DB	1009 CCAATCGAAAGCTGTCTCAACCATCTCTTCCAGAGTGGCCCAAAATGTAAAGACACAGCTGC	1066		
QY	842 TCAATTAGATCTTATCATGATCTTGTGAGCCAGCAGACAGCGCTTTCGAAATTCAGGGAG	901		
DB	1069 TCAATTAGATCTTATCATGATCTTGTGAGCCCAACAAATGATCCCGCTTCGAAATTCAGGGAG	1128		
QY	902 TGGGCAATATCAGCTATGAGATTTCTTACTGAGAGCTTCTGTGGGACAGTCTCAACTCCA	961		
DB	1129 TGGGCAATATCAGCTATGAGATTTCTTACTGAGAGCTTCTGTGGGACAGTCTCAACTCCAG	1188		
QY	962 CTGCACTACCTGGGAGAGGCAACAATGGGAGTTCAGAGATGAAGACCCGATGAAGTGGC	1021		
DB	1189 CTGCACTACCTGGGAGAGGCAACAAGGGAGTTCAGAGATGAAGATCCCGAGAGAGTGGC	1248		
QY	1022 TGGCGGTTGGGGAGAGAGAAAGCAACCTTAACATGAACATATGACAAACTCAGCGCTGC	1081		
DB	1249 CCGCGGCTGGGGAGAGAGGAAAGCAACCAACATGAACATATGAAAGTCAAGCTCAGCGCGC	1308		

QY	1082	ACTGGTACTCTCTATGACCAAAAATATATATAGCTAAAGTCAATGCTGAACAGCTATAGCCCA	1141
Db	1309	CTTCGGTACTACTATGATGACAAAGATATCATACCAAGGCTATGGGAAAGCGGTACGCCCA	1368
QY	1142	CAAAATTTGATTTCCACAGGAAATGCTCAGGCCCTCCAGGCTCACCCCTCCAGAAATCATTCAT	1201
Db	1369	CAAGTTGCAGCTTCCACGGGATGCGCCAGGCCCTCCAGGCCACCCCGCGAGTCACTCTCT	1428
QY	1202	GTACAAATACCATATAGACCTTCCTTACATGATGATTCTTACATGCAACACCCCAAGAGAT	1261
Db	1429	GTACAAAGTACCCCTCAGACCTCCCTCATAGTGGCTCTCATACAGCCCAACCAAGAT	1488
QY	1262	GAACCTTGTAGCTCCCAACCCCTCCGCTTTGGCCGCAACCTCATCAGCTTTTGGCTGC	1321
Db	1489	GAACCTTGTGGGGGCCCAACCTCCAGGCCCTCCCTCCGATCATTTTTCAGTTTTTGGCTGC	1548
QY	1322	CCCTAATCATACTGGAATTCACCAACTGGAGGACATCAACCCAAATACAGGCTCCAGC	1381
Db	1549	CCCAACACCATCTACTGGAATTCACCAACTGGGGGTATATACCCCAACACTAGGCTCCAC	1608
QY	1382	TGCTCATATGCTTCGCCATCTTGGGACCTACTACTCA	1418
Db	1609	CAGCCATATGCTTCCTCATCTGGGCACTTACTACTCA	1645
RESULT 7			
LOCUS	XLAJ4126	2012 bp	mRNA linear VRT 21-JAN-2000
DEFINITION	Xenopus laevis erg gene (erg_E).		
ACCESSION	AJ224126		
VERSION	AJ224126.1 GI:5420047		
KEYWORDS	ERG gene; transcription factor.		
SOURCE	Xenopus laevis.		
ORGANISM	Xenopus laevis		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae; Xenopodinae; Xenopus.		
AUTHORS	1 (bases 1 to 2012)		
TITLE	Baltzinger, M., Mager-Heckel, A.M. and Remy, P.		
JOURNAL	XI erg: expression pattern and overexpression during development		
MEDLINE	Dev. Dyn. 216 (4-5), 420-433 (1999)		
PUBMED	20099678		
REFERENCE	10633861		
AUTHORS	2 (bases 1 to 2012)		
TITLE	Baltzinger, M.		
JOURNAL	Direct Submission		
FEATURES	Submitted (24-FEB-1998) Baltzinger M., UPR9005, NMDCD, Centre National de la Recherche Scientifique, 15, RUE Rene Descartes, 67084, FRANCE		
SOURCE	Location/Qualifiers		
gene	1..2012		
CDS	/organism="Xenopus laevis"		
	/db_xref="taxon:8353"		
	/clone="lambda ZAP11-erg E"		
	/clone_lib="lambda ZAP11"		
	/dev_stage="stage 30 embryo"		
	225..1595		
	/gene="erg"		
	225..1595		
	/gene="erg"		
	/function="Transcription factor"		
	/codon_start=1		
	/protein_id="CABA6567.1"		
	/db_xref="GI:5420048"		
	/db_xref="SPTRKBL:O9W629"		
	/translation="MIQTVDPSPAHIKREALSVSEDSLEFCYGTPLHTKTEMTASS		
	SSDVGQSKMSPRVQDDWLSQPSRTYIKKECNPSOVNGSRSSPDSCISGSKMGSG		
	GSNDVGNWYSGYIEEKHSIPNMTNRRVTVPADPLMSDHYRQLFWAIKRYGILP		
	DVYVLQNNDIGKELCKRTKDFRLRPLSPYADLLSLHLIRKRGASFIFPNSVYQ		
	DANRILSRQDLSTEPSRSAMTNHPAPSAASPSTVPEKTEPDRQOLDYQILGLPTP		
	SSRLANSGGQIQIMQFLLELLSDSSNSNCTWGTNGEFTMDPDEYARWGERKSK		

		PNNYDKLSRLRYDYKDKINTKXGHRVAKPFPHJIAOLQHPRESTYKXPSLEI	
		PYMSYHAHPQKMFVAPHPRLALVTSSSPFAANWATNBPSTGSIYNTRLPASHMS	
		HLGITYY"	
BASE COUNT	634 a 479 c 395 g 504 t		
ORIGIN			
Query Match	61.5%; Score 889.6; DB 5; Length 2012;		
Best Local Similarity	79.7%; Pred. No. 7.6e-259;		
Matches 1077; Conservative	0; Mismatches 269; Indels 6; Gaps 2;		
OY	74 TATTAAAGGAACATTATCAGTGGTGAAGTAAGACAGCTCTGTGTTGAGTGCCTACGG	133	
Db	257 TATCAAGAGAGCTCTTCAGTGTGAGTGAAGATGACGTGTTATTGCGATGGACCTACGG	316	
OY	134 ATGCCCCCACTTCGAAAAGACAGAAATAGACGCTCTCTTCCAGTGAATATGGCAAC	193	
Db	317 AACGCCACATCTTACTTAAAGCGAGATACCGCATCTTCTCCAGTACTATGGCCAAAC	376	
OY	194 ATCAAAAGTGAAGCCCGCGCTGCCACAGAGACTGGTTATACAGCCCCCGCGCAGAGT	253	
Db	377 CTCGAAATAGACTCCAGCTGTGCCCTCAACAGACTGGCTCTCACAGCCTCCATCCAGGT	436	
OY	254 TACACTTAAGATGAGTGTAAACCCAAACAGGTTAATGGTCAAGAGTAATCACTGATGA	313	
Db	427 GACCATCAAGATGGAATGCAACCCCTGACAGTCAATGATGCAAGAGACTCCCGGATGA	496	
OY	314 CTCGACGCGTGGCAAAAGAGGGAAATGGTTTGCAGTTTCAGACAAATGTTGGAGTACCTA	373	
Db	497 CTCGAGTATAGAAAGAGACCAAAATGGGTGGAGGGTGGGATATGTTGGAATGAATTA	556	
OY	374 TGGAAAGTNCATGGAAGAGAAAGCATATTCGCGCTCCCAATATGACAAACCATTGAACGAG	433	
Db	557 TGGCAGCTACATAGAAAGAAAACACATTTCTCCCCCAAACTGACACACCAATGAGAGAG	616	
OY	434 AGTTATTTGGCAGAGAGATCCCTAGCTTATATGAGCAGACACCATGATGACGGCTGGA	493	
Db	617 AGTGATTTACTGCTGTGACCTTACTTTGTGAGACACATATGTGACAGAGTGGCTTGA	676	
OY	494 GTGGCAGTGAAGAGTATGGTCTTCCAGAGCTGGACATCTTGTGTTCCAGAACATTGA	553	
Db	677 GTGGCCATTAAGAAATACGGGCTTCCAGATGTGATGTGCTCGTTCGTTCCAAACATGGA	736	
OY	554 TGGGAAAGAGTTGTGTAAATATGACCAAGATGACTTCCAGAGACTCACGCCAGCTATTA	613	
Db	737 CGGCAAGAAATATGCAAGATGACCAAAAGAAATTTCCATGAGACTCACGCCAGCTACAA	796	
OY	614 CGCAGATATCCCGCTGTACACCTTACACTTACTCTGAGAGAGAGGACACTTTATATTT	673	
Db	797 TGCATATATCTCTCTCTGTGCATCTTCACTACTCTCAGAAAGAGTGGCTCTTTTATTTT	856	
OY	674 TTCAAATACATCAGTTTATCCAGAGCAAGCAGCAAGAAATTAACAACAGCCAGATTTACC	733	
Db	857 CCCAAACTCATCCGTTTATCAAGAGCAACCAAAAGAAATCCCGACAGCAAGATTTATTC	916	
OY	734 TTATGAGCAAGCGAGGAATCAGGCTGGACGATGCACAGCCATCCCACTCACTCAAAAGC	793	
Db	917 ATATGAGCCCTCAAGAGATCTGCTTGGACAATCA---TCTGTGACCCACCATCAAAAAGC	973	
OY	794 TACCACACCATCATCTTCAACAGTGCACCAAAACAGAAACACAGCGTCTCGATTAGATCC	853	
Db	974 ATCTCAAC---CATTTACCAACAGTTCCCAAAACAGAAACCCAGCCACCTGATATCC	1030	
OY	854 TTATCAGATCTTGTGGACCGACGACGCGCTTGTGCAATCCAGGAGATGTGGCAGATACA	913	
Db	1031 TTATCAGATCTTGTGGCCCAACAAGACGCCGACTTGCAAATTCAGGAAGTGGGCAATTCA	1090	
OY	914 GCTATGCAAGTCTCTACTGAGCTTCTGTGCGACAGCTCCAACTCCAACTGCATCACCTG	973	
Db	1091 GCTGTGGCAATTTCTACTGGAAGCTGTTTGGATATGTTCCAACTCCAACTCATTTACTTG	1150	
OY	974 GGAAGGCAAAATGGGAGTTCAAGATGACAGACCCGTGATGAAGTGGCTGGCGTTGGGG	1033	
Db	1151 GGAAGGAACCATATGGAGATTTAAGATATCCGATCTCTGATTAAGTGGCAAGACCTTGGG	1210	


```

QY 655 --GAGAGCCACTTTATTTTCCAAATACATCATGTTTACCAGAACGCAAGAAT 712
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1019 AGGAGGTCCTCTTTATTTTCCAAACTCATCGTTTATATCAAGACCAAGCAAGAAAT 1078
QY 713 AACAAAGGCCAATTTACTTATGAGCAAGGAGATCAGCGTGGACGATCAGC 772
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1079 CCCCAGACGAGAAATTTATATATGAGCCCTCAAGGAGATCTCTGGCAACAAATCAT - 1136
QY 773 CCATCCCACTGATCAAAAGGTACCAACATCATCTTCAACAGTGGCCCAAAACAGAA 832
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1137 -CTGCAACACCATCAAAAGCATCTCAAC---CATCTCCACAGTTCCCAAAACAGAA 1192
QY 833 CCAGCGTCTCAGTTAGTCTTATCAGATTCTTGGACGACGACCGCTCTGCAA 892
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1193 CCCCAGGCGACAGCTAGATCTTATCAGATTCTTGGGCAACAGACGCCACTTGCAAA 1252
QY 893 TCCAGGAGTGGGAGATACGATATGCAATCTTCTACTGAGAGCTTCTGTGGACAGTC 952
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1253 TCCAGGAAGTGGGAGATCTGCAATCTTCTTACTGGAAGCTTCTTGGATAGTTC 1312
QY 953 CAATCCCACTGATCAAAAGGTACCAACATCATCTTCAACAGTGGCCCAAAACAGAA 1012
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1313 CAATCCCACTGATCAAAAGGTACCAACATCATCTTCAACAGTGGCCCAAAACAGAA 1372
QY 1013 TGAATGCTCTGGGCTGGGAGAGAGAGAAACCAACCTAATGATGATGACAAAT 1072
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1373 TGAATGCTCTGGGCTGGGAGAGAGAGAGAAACCAACCAATGATGATGACAAAGCT 1432
QY 1073 CAGCCGTCGACTTGGCTACTATGACAAATATTTATGATGATGATGATGATGAT 1132
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1433 CAGCCGTCGACTTGGCTACTATGACAAATATTTATGATGATGATGATGATGATGAT 1492
QY 1133 CTATGCTTCAAAATTTGATTTCCAGGAATGCTCAGGAGCTTCCAGGCTTCCAGGCT 1192
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1493 CTATGCTTCAAAATTTGATTTTCAATGAGATTTGCTCAGGCTTCAACCTTCACTCAGA 1552
QY 1193 ATCATCATGATGATCAAAATCCCATGAGACCTTCCATGATGATGATGATGATGATGAT 1252
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1553 ATCAACCATGATGATCAAAATCCCATGAGACCTTCCATGATGATGATGATGATGATGAT 1612
QY 1253 CCAGAAATGATGATTTAGTCTCCCATCCCTGCTTGGCTTGGCTTGGCTTGGCTTGGCT 1312
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1613 ACAGAAATGATGATTTAGTCTCCCATCCCTGCTTGGCTTGGCTTGGCTTGGCTTGGCT 1672
QY 1313 TTTTGTGCTCCCTAATCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1372
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1673 CTATGAGAGCCCTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1732
QY 1373 GCTGCGAGCTGCTATATGCTTCCCATCTTGGCACTTACTACTAGTGGGGA 1425
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1733 GCTGCGAGCTGCTATATGCTTCCCATCTTGGCACTTACTACTAGTAAATCAAGA 1785

```

```

RESULT 12
HUMERG11 3126 bp mRNA linear PRI 27-Apr-1993
LOCUS Human erg protein (ets-related gene) mRNA, complete cds.
DEFINITION M21535.1 GI:182182
ACCESSION M21535.1 GI:182182
VERSION 1 of 2
KEYWORDS erg protein.
SEGMENT 1 of 2
SOURCE Human, cell line COLO 320, cDNA to mRNA, lambda-7.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 3126)
AUTHORS Reddy,E.S., Rao,V.N. and Papas,T.S.
TITLE The erg gene: a human gene related to the ets oncogene
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 84 (17), 6131-6135 (1987)
MEDLINE 87317608
PUBMED 3476934
FEATURES
location/Qualifiers

```

```

source 1. 3126
/organism="Homo sapiens"
/db_xref="taxon:9606"
195..1286
/note="erg1 protein"
/codon_start=1
/protein_id="AA35811.1"
/translaton="MVGSPDVTGNYGSEKHKMPPNNMTTNERIVIPADPTLMST
DHVROWLEMAKVEYGLPDVILIFONIDGKELKMPDDBORLTPSYNADILSHLY
LRREPLHLTSDVDYKRALONSPRLMARNDLPEPRBSAWGHGPTQSKAOPS
PSTVKTEDORPOLDPYQILGPTSSRLANSGOIQLMQLELSDSSNSCTTWBG
TNGEFTKDPDEVARWGERKSRPNMYDRLSLRLRYDKNTKTKYGRKAYKDFP
HGIAOALQPHRPPSSLYKYPSTLRYMSYAHQKMFVAPHPALPVTSSSFEAPN
PYNSPGGIYPTVRLPTSHMPSHLGYY"
BASE COUNT 928 a 732 c 725 g 741 t
ORIGIN 1 bp upstream from EcoRI site.
Query Match 52.5%; Score 760.2; DB 9; Length 3126;
Best Local Similarity 80.7%; Pred. No. 1.8e-219;
Matches 917; Conservative 0; Mismatches 208; Indels 12; Gaps 2;
QY 294 TCAAGAAATTCACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 353
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 150 TCAAGAAATTCCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 209
QY 354 GACATGCTTGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 413
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 210 GACACCGTTGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 269
QY 414 ATGACCAACCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 473
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 270 ATGACCAACCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 329
QY 474 CATGTAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 533
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 330 CATGTAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 389
QY 534 TTTGTTGTTCCAGAACTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 593
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 390 TTTGTTATTTCCAGAACTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 449
QY 594 AGACTGAGGCGGAGCTATTAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 653
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 450 AGGCTACACCCCACTCAACAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 509
QY 654 A-----GAGAGGCACTTTATTTTCCAAATATGATGATGATGATGATGATGATGATGAT 704
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 510 ACTCTCTTCCCACTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 569
QY 705 CAAGAAATTAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 764
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 570 TTAATGATGATGATTAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 629
QY 765 AGTACAGACCATCCCA---CTCAGTCAAAAGCTAACCAACCATCATCTTCAACAGTGGCC 821
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 630 GGTGACGCGCACACCCACGCGCCCACTGAAAGCTGCTCAACCATCTCTTCCACAGTGGCC 689
QY 822 AAAACAGAAAGCAAGCGCTCCTCACTTATGATGATGATGATGATGATGATGATGATGATGAT 881
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 690 AAAACAGAAAGCAAGCGCTCCTCACTTATGATGATGATGATGATGATGATGATGATGATGAT 749
QY 882 CGTCTTCAAAATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 941
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 750 CGCTTCAAAATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 809
QY 942 TCGGACAGCTCCAACTCCAACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1001
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 810 TCGGACAGCTCCAACTCCAACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 869
QY 1002 ACAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1061
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 870 ACAGATCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 929

```

BASE COUNT	377 a	439 c	349 g	271 t	ORIGIN
Query Match	39.5%	Score 571.2;	DB 5;	Length 1436;	
Best Local Similarity	65.9%	Pred. No. 5,4e-162;			
Matches 901;	Conservative 0;	Mismatches 443;	Indels 24;	Gaps 4;	
QY	63	ATGGCAAGCACTATTAAAGGAAGCATTTATCACTGGTGAAGTGAAGCAACCACTCTGTTTGAG	122		SGSIQIOWMFLLELSDSANNACJLWETSENGEEMFKMTDDEYARMRGKESSKPNMNYDKL
DB	45	ATGGACGGCAACTATTAAAGGAGCGCGTGTCACTGGTGAAGTGAAGCAACCACTCTGTTTCAG	104		SRALRYXYDKNIMTKVHGKRYAFYEDPHGIAOLAOPHTPEVAKYSESLPYVYSYRA
QY	123	TGTGCTTACGATATGCCCCCACTTGCA---AAGACAGAAATGACAGCCTCTCTTCCAGT	179		HOOKNVFSPHPSPMPTSSNFGEPTTPWYSSPPGGIYIRPNPSVRHANSHPHSLSY
DB	105	CCTCATATGCGCCCTGCTGCTCTTATCCCAAGACAGACATGACGTCTCCGGACATCAG	164		
QY	180	GAATATGGGCAAAACATCAAGATGAGCCCGCGCTTCCCAAGCAAGACTGGTTATACACAG	239		
DB	165	GATTATGCGCCAGACGACAAATCAACCCCATTCCTCCCTCAGCAAGAGTGGATTAAATCAG	224		
QY	240	CCCCCGGCCAAGTTACACTTAAAGTGAAGTGTAAACCAACCAAGTTAATGGGCTCAAGG	299		
DB	225	CC-----Ggtgcggggtgaacgtcaaggagagatgacacatcacaacggatccaga	275		
QY	300	AATTCACTGATGACTGACGCGTGGCAAAAGAGGAAATGGTTAGCATGTTACAGCAT	359		
DB	276	GAGTGCCTGGTGGACTGACGCGTGGGAAATGACATATAATATGGTGGGGGACCGAAACG	335		
QY	360	GTTGGGATGAACATATGAGAGCTACATGGAAGAAGCATATTCGCGCTCCAAATATGACA	419		
DB	336	TCTCAGATGAACATATCTAGCTACATGATAGAAATGTGCTCCGCTCCCAACATGACG	395		
QY	420	ACCAATGAACAAGATGATTTGTGCACACAGATCCCTACGTTATGGAGACACAGACATGTA	479		
DB	396	ACCAATGAGAGAGAGATGATGCTGCTCCCGCAACCCCGTCTCTGTGTGCTCCGGATCAGTG	455		
QY	480	CGGAGTGGCTTGGAGTGGGACAGTGAAGGATATGGTCTTCCAGACGTTGACATCTTGTG	539		
DB	456	CGGAGTGGTATGAGATCGGGCCATTAAAGGATACGGTCTTCAGAGATGACACCGGCATG	515		
QY	540	TTCAGAAACATTTGATGGGAAGAGTGTGTAAATGACCAAAAGATGACTTCCACAGATC	599		
DB	516	TTTCACACGACAGATGGAAGAAAGACTCTGCAAGATGACAAAGACGACTTCTTCAGATCTC	575		
QY	600	ACGCGGAGCTTTAAGCGCAGATATCTCTGTGCACACCTACACTACCTACAGAGAGAGAGA	659		
DB	576	ACCAAGCGTTTAAACACGAGAGTCTCTCTCTCACATCTCAATTTACTCAGGAAAGTAG-	634		
QY	660	GCCACTTTATTTTCCAAATACATCACTGTTTACCAGAGACGACGCAAAAGATTAACA	719		
DB	635	--CTCATCGATATATATACACACGCACTCTACGCTGACAGTCCCTCGCTGGCTGCA	692		
QY	720	AGGCGAGATTTCACCTTATGAGCAAGCGGAGAGATCAGGTGGACGAGTCAAGGCATCCC	779		
DB	693	AAAGACGATGCGTCTTATGATGCTGTACGGAGACGGGCTGTGTGAACAAATGACACAGC	752		
QY	780	ACTGAGTCAAAAGGTATACCAACATCACTTCAACAGTGGCCCAAAAGAAAGACAGCGT	839		
DB	753	GGAAGAGCTCTCCACACAGTGTGTCTCAGAGCGGTGTCAAAGAACCGGATCAGCGTGT	812		
QY	840	CCTCAGTTAAGTCTCTTATCAGATCTTGGACCGACGACAGCAGCGCTCTTGCAAAATCCAGG	899		
DB	813	CCTGAGCGAGTCTCTTATCAGATATTAAGGACCCACACAGCACTGCACTTGTAAACCAAGC	872		
QY	900	AGTGGGAGATACAGATATGACGACTTCCCTACTGTGAGAGCTTGTGGACAGCTCCAACTCC	959		
DB	873	TCAAGTCAAGATCCAGCTGTGGCACTTCTGTGTGAAGCTCTGTCCAGACGGCCAAACGC	932		
QY	960	AACCTGATCACTTGGGAGGCGCAAAATGGGAGTTCAAGATGACAGACCTGATGAAGTG	1019		

Db 933 GCGTCATCACTGGAGAGAACCAAGAGAGTTCAAGATGACGAGACCCAGAGAGT 992
QY 1020 GCTCGGCGTTGGGAGAGAGAAAACAACTAATGATGACTATGACAACTGAGCGT 1079
Db 993 GCACGGCGCTGGGGGCAAGCAAGAACCAACCAATGAACTGACAAAGCTGAGCCGA 1052
QY 1080 GCACCTGGCTACACTATGACAAATAATATATGACTAAATGCTGATAACGTATGCC 1139
Db 1053 GCGCTGGCTACTACTAGCAAGAACATCATGACCAAGATGACGAGGAGCGGTACGCC 1112
QY 1140 TACAAATTTATTTCCACGGAATGCTCAGGCGCTCCAGCGCTCCAGCAATATATCC 1199
Db 1113 TACAAAGTTCACCTCCACGCGATCGCCAGCGCTCCAGCAATATATCCAGCTCAC 1172
QY 1200 ATGTACAATACCCATCAGACCTCCCTCATGATGATGATGATGATGATGATGATG 1259
Db 1173 ATGTACAATACCCATCAGACCTCCCTCATGATGATGATGATGATGATGATGATG 1232
QY 1260 ATGAACTTTGATGCTCCCATCCCTGCTTGGCCGCTGATGATGATGATGATGATG 1319
Db 1233 GTCAACTTCTGCTCTCCACCGCTGCTCAATGCTGCTCAATGCTGCTCAATGCTG 1292
QY 1320 GCGCTATATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1379
Db 1293 CCACAAACCCATATCTGAGAGCTCCCGACAGAGAGATTTTACCCCAATCCAGCGT 1352
QY 1380 -----GCTGCTATATGCTTCCCATCTTGGCAGCTGATGATGATGATGATG 1418
Db 1353 CGCATGCAATTCACAGCTGCTTCACTGATGATGATGATGATGATGATGATGATG 1400

RESULT 14
MMFL1 1729 bp mRNA linear ROD 02-SEP-1991
LOCUS Mouse Fl1-1 mRNA for retroviral integration site.
DEFINITION X59421
ACCESSION X59421
VERSION X59421.1 GI:50974
KEYWORDS c-ets-1 gene; ets gene family; fl1-1 gene; integration site.
SOURCE Mus musculus.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1729)
AUTHORS Letwin,K.
TITLE Direct Submission
JOURNAL Submitted (08-MAY-1991) K. Letwin, Samuel Lunenfeld Res Inst at Mount Sinai Hospital, Div of Mol and Developmental Biol, 600 University Avenue, Toronto Ontario M5G 1X5, CANADA
2 (bases 1 to 1729)
Ben-David,Y., Giddens,E.B., Letwin,K. and Bernstein,A.
Erythroleukemia induction by Friend murine leukemia virus: insituational activation of a new member of the ets gene family, Fl1-1, closely linked to c-ets-1
JOURNAL Genes Dev. 5 (6), 908-918 (1991)
MEDLINE 91257578
PUBMED 2044959

FEATURES
source
Location/Qualifiers
1..1729
/organism="Mus musculus"
/strain="balb/c"
/db_xref="taxon:10090"
/cell_line="C87"
/cell_type="Friend erythroleukemia"
/germline
1..1729
/gene="Fl1-1"
1..1729
/gene="Fl1-1"
/evidence="experimental"
245..1603
/gene="Fl1-1"
/codon_start=1
/product="retrovirus integration site"

CDS
gene
mRNA
CDS

/protein_id="CAA42055.1"
/db_xref="GI:50975"
/db_xref="MD:MG1:95534"
/db_xref="SWISS-PROT:P26323"
/translation="MDGTEKALSVSDSDSLFSAVGAHLPRKADMTASGSDYGO
PKHINLPDQEWINQPVRYNRYREYDHNHNSRESVDCSVSKCNLTGCGEAPNRY
NSTMDENKNDPPNNTTNERVIVPADPILMTQDEHRYOMLEAIKEKYGIMEDITSFQ
NMDKEICRNKEDFLATSAVTEVLVLSLSTRSSLIATNTSHGDSRLNVE
DPSYDVRGRGAMNNNSGLNKSPLDGSQTMKNTQRPDPYQILGPTSSRLANP
GSGQIQLMQFLLELSDSANASCITWEGTGEFKMTDPDEARMEGRSKPMYNDK
LSRALRYVDKNTMTVHGKRYRYKFDPIGIALDLPHPETSTVPSYPSISYH
AHQOKXNFPVSHSSMPVTSSTPFGAASQYWNISPTAGIYRPNPVRHPTNHYSHLS
Y"

BASE COUNT 466 a 484 c 436 g 343 t

ORIGIN

Query Match 39.4%; Score 570.2; DB 10; Length 1729;
Best Local Similarity 66.8%; Pred. No. 1.1e-161;
Matches 917; Conservative 0; Mismatches 423; Indels 33; Gaps 6;

QY 63 ATGGCAAGCACTATTAAAGAGCAATTAATCAATGATGATGATGATGATGATGATGATG 122
Db 245 ATGAGAGGAGCACTATTAAAGAGGCTCTGCTGATGATGATGATGATGATGATGATG 304
QY 123 TGTGCTACGAGATC---CCCCACCTTGCAAGACAGCAATATGACAGCTCTCTTCACT 179
Db 305 TCACATATACGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 364
QY 180 GAATATGAGGCAATCAATGAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 239
Db 365 GACTAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 424
QY 240 CCCCCGCGAGAGTTTACATTAATGATGATGATGATGATGATGATGATGATGATGATG 299
Db 425 CC-----AGTGAAGTCAATGTCAAGGAGGAGTATGACCAATGATGATGATGATG 475
QY 300 AATTCACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 359
Db 476 GAGTCTCGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 535
QY 360 GTTGGATGAGCACTATGAGGATGATGATGATGATGATGATGATGATGATGATGATG 416
Db 536 AACCCATGACATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 595
QY 417 ACAACCAATGACAGCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 476
Db 596 ACACCAAGAGAGGAGGAGGATGATGATGATGATGATGATGATGATGATGATGATGAT 655
QY 477 GTACGCAAGTGGCTGAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 536
Db 656 GTTCGACAGTGGCTGAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 715
QY 537 TTGTTCCAGAACTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 596
Db 716 TTCTTCAGAACTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 775
QY 597 CTCAGCGGAGCTTAAAGCAGATATCTCTGACACCTACCTACCTACCTACCTACCT 656
Db 776 GCGACCTCGGCTTAAAGCAGAGTGTGTGTGACACCTACCTACCTACCTACCTACCT 834
QY 657 GGAGCACTTTTATTTTCCAAATATCATGATGATGATGATGATGATGATGATGATG 716
Db 835 -----TTCACTGCTGAGCTTAAACCAACCTCCATACAGCAGTCTCAGCTGAAT 889
QY 717 ACAAGGCAATTTACCTTATGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 773
Db 890 GTCAAGGAGACCTTCTTATGACTGTGACAGGAGGAGGAGGAGGAGGAGGAGGAGG 949
QY 774 CATCCACATGATCAAAAGTACCAACCATATCTTCAACAGTGCACCAAAACAGAGAG 833
Db 950 TGTGGCTCAAAAGATCTCTCTGAGAGATCAAGACCATGAGGAGGAGGAGGAGGAG 1009
QY 834 CAGCGTCTCAGTATGATCTTATCAGATTTCTTGAGACGACGACGACGACGCTTGA 893

Db	1186	TATCTTACAAATTTAGCTCCAGGCACTTGGCCAGGCTCTGCACCCACATCCACGGAG	1245
Qy	1194	TCATCCATGTACAAATACCCTATAGACCTCGCCCTCATGAGTTCCTACCATGACACGCC	1253
Db	1246	TGCGTCATGTACAAAGTACCCCTTCTGCACATCTCTCATATGCCCTTCTTACATAGCCCAACAG	1305
Qy	1254	CAGAAATGAACTTTGTAGTCCCATCCCTGCTTTTGGCCGTAACCTATCCAGCTTT	1313
Db	1306	CAGAGGTGAACCTTGTGCCCTCCCTCCCATCTCCATGCGCTGTACCTCTCCAGCTTC	1365
Qy	1314	TTTGTCTCCCTATATCATCTGTAATTCACAACTGTGAGGATCTATACCCCAATACGA - -	1371
Db	1366	TTTGGACCCGCATACAAATACTGACCTCCGCCACAGGGGGAAATCTACCCCAACCCCAAC	1425
Qy	1372	-----GGGTGGCAGGTGTGCTATATGCGTCCCATCTTGGCACTACTACTA	1417
Db	1426	GTCCCCCGGCATCTTACACCCACGTCCTTACACTTATGGAGGTACTACTA	1478

Search completed: November 9, 2002, 11:35:56
Job time : 3957.96 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 9, 2002, 09:08:56 ; Search time 414.84 Seconds
(without alignments)
7855.185 Million cell updates/sec

Title: us-09-902-772-1

Perfect score: 1447
Sequence: 1 gaattccgcgaacgaataat.....gaaagaagcgcacagaaaa 1447

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

N_Geneseq_101002:*

- 1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
- 2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
- 3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
- 4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
- 5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
- 6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
- 7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
- 8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
- 9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
- 10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
- 11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
- 12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
- 13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
- 14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
- 15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
- 16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
- 17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
- 18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
- 19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
- 20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
- 21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
- 22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
- 23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
- 24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1447	100.0	1447	20	AAAX26551	DNA encoding chick
2	1356	93.7	1528	20	AAAX26552	DNA encoding chick
3	563.8	39.0	2938	14	AAO50644	Human Hum-Fli-1 ge
4	562.2	38.9	2957	24	ABR84139	Human CDNA differe
5	544.6	37.6	2954	22	AAH02915	Human shear stress
6	387	26.7	567	22	ABA48124	Human breast cell
7	387	26.7	567	22	ABA66003	Human foetal liver
8	387	26.7	567	22	ABA33090	Probe #11556 for g
9	387	26.7	567	22	AAK14426	Human brain expres

C 10	387	26.7	567	22	AAK40160	Human bone marrow
C 11	387	26.7	567	22	AAI20932	Probe #10865 for g
C 12	387	26.7	567	22	AAI46176	Probe #14862 used t
C 13	387	26.7	567	22	AAI06643	Probe #6634 used t
C 14	387	26.7	567	22	ABR514215	Human genome-deriv
C 15	354.6	24.5	473	22	ABR50193	Human breast cell
C 16	354.6	24.5	473	22	ABR68128	Human foetal liver
C 17	354.6	24.5	473	22	ABR35152	Probe #13618 for g
C 18	354.6	24.5	473	22	AAK16512	Human bone marrow
C 19	354.6	24.5	473	22	AAK42285	Human bone marrow
C 20	354.6	24.5	473	22	AAI48339	Probe #17025 used
C 21	354.6	24.5	473	22	AAI08696	Probe #8687 used t
C 22	325	22.5	533	21	AAAC01342	Human secreted pro
C 23	320.4	22.1	1890	14	AAO50662	Human Hum-Fli-1 ge
C 24	276.6	19.1	549	24	ABR44114	CDNA #54 encoding
C 25	255	17.6	420	22	ABR45056	Human foetal liver
C 26	255	17.6	420	22	ABR55529	Human breast cell
C 27	255	17.6	420	22	ABR25239	Probe #3705 for ge
C 28	255	17.6	420	22	AAK03760	Human bone marrow
C 29	255	17.6	420	22	AAK29224	Human bone marrow
C 30	255	17.6	420	22	AAI35180	Probe #3866 used t
C 31	255	17.6	420	22	AAI03692	Probe #3683 used t
C 32	255	17.6	454	22	ABR42995	Human breast cell
C 33	255	17.6	454	22	ABR53411	Human foetal liver
C 34	255	17.6	454	22	ABR23187	Probe #1653 for ge
C 35	255	17.6	454	22	AAK01681	Human bone marrow
C 36	255	17.6	454	22	AAK27129	Human bone marrow
C 37	255	17.6	454	22	AAI11718	Probe #1651 for ge
C 38	255	17.6	454	22	AAI33025	Probe #1711 used t
C 39	255	17.6	454	22	AAI01646	Probe #1637 used t
C 40	255	17.6	454	24	ABR01685	Human genome-deriv
C 41	235.6	16.3	318	21	AAAC1875	Human secreted pro
C 42	180	12.4	1752	22	AAI33672	CDNA encoding Ralt
C 43	180	12.4	1752	24	AAI38753	Rat Lambda73 CDNA
C 44	170.6	11.8	899	23	ABR12097	Drosophila melanog
C 45	167.8	11.6	553	22	ABR61050	Human foetal liver

ALIGNMENTS

RESULT 1	AAAX26551	standard; DNA; 1447 BP.
ID	AAAX26551	
XX	AAAX26551:	
AC	14-JUN-1999	(first entry)
XX		
XX	DNA encoding chicken C-11 protein.	
DE		
XX	Chicken: C-11 protein; cell calcification inhibiting activity;	
KW	cell calcification inhibiting agent; c-erb protein; arthritis deformans;	
KW	ossification; spinal column ligament; ss.	
XX		
OS	Gallus sp.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	63..1418
XX		/*tag= a
XX	JP11075871-A.	
PN		
XX	23-MAR-1999.	
PD		
XX	29-MAY-1998:	98JP-0166076.
PE		
XX	20-JUN-1997:	97US-0050297.
PR	18-JUN-1997:	97US-0878177.
XX		
PA	(CHUS) CHUGAI PHARM CO LTD.	
XX	(OYPE-) UNITV PENNSYLVANIA.	
XX		

DR WPI: 1999-257708/22.
 P-PSDB: AA01520.
 XX An active protein for inhibiting cell calcification - useful for
 PT measuring the calcification of a cell, for diagnosing arthritis
 PT deformans or ossification of spinal column ligament
 PS Disclosure: Page 7-8; 15pp; Japanese.
 XX
 CC The present sequence encodes a chicken C-11 protein which has cell
 CC calcification inhibiting activity. The specification also describes
 CC a cell calcification inhibiting agent containing c-ery protein
 CC (AA01521). The proteins are used for measuring the calcification of a
 CC cell, for diagnosing arthritis deformans or ossification of spinal column
 CC ligament.
 CC
 XX
 SQ Sequence 1447 BP; 440 A; 374 C; 317 G; 316 T; 0 other;

Query Match 100.0%; Score 1447; DB 20; Length 1447;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1447; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAATCCGGGAGCAATTAATTTATTTAGCAATTTATAGCAATTAATCTGTGATCACA 60
 DB 1 GAATCCGGGAGCAATTAATTTATTTAGCAATTTATAGCAATTAATCTGTGATCACA 60
 QY 61 TTATGGCAAGCACTATTAAAGAAAGCATTTACAGTGTGAGTGAAGCAAGCTCTGTTTG 120
 DB 61 TTATGGCAAGCACTATTAAAGAAAGCATTTACAGTGTGAGTGAAGCAAGCTCTGTTTG 120
 QY 121 AGGTGCTCAGAGATGCGCCACCTTGCAAAAGACAGAAATGACAGCTCTCTCCAGAG 180
 DB 121 AGGTGCTCAGAGATGCGCCACCTTGCAAAAGACAGAAATGACAGCTCTCTCCAGAG 180
 QY 181 AATATGGGAAACATCAAAAGATGAGCCCGCGCTTCCCGAGCAGACTGGTTATCAGAC 240
 DB 181 AATATGGGAAACATCAAAAGATGAGCCCGCGCTTCCCGAGCAGACTGGTTATCAGAC 240
 QY 241 CCCCAGCCAGAGTTACCAATTAGATGAGTGAACCCAAACAGGTTAATGGTCAAGGA 300
 DB 241 CCCCAGCCAGAGTTACCAATTAGATGAGTGAACCCAAACAGGTTAATGGTCAAGGA 300
 QY 301 AATCACCCTATGCTCAGCTGCGCAAAAGAGGAAATGGTTACAGTTCAGACAAAG 360
 DB 301 AATCACCCTATGCTCAGCTGCGCAAAAGAGGAAATGGTTACAGTTCAGACAAAG 360
 QY 361 TTGGGATGAACCTATGAGAGCTACATGGAAGAGATATTCGGCTCCAAATATGACAA 420
 DB 361 TTGGGATGAACCTATGAGAGCTACATGGAAGAGATATTCGGCTCCAAATATGACAA 420
 QY 421 CCAATGACGAAAGATTTATGTGCGACAGATCTTACGTTATGGAGACAGACCATGTAC 480
 DB 421 CCAATGACGAAAGATTTATGTGCGACAGATCTTACGTTATGGAGACAGACCATGTAC 480
 QY 481 GGGAGTGGCTGAGTGGGAGTGAAGAGATGAGTCTTCAGACGCTGACATCTTGTGT 540
 DB 481 GGGAGTGGCTGAGTGGGAGTGAAGAGATGAGTCTTCAGACGCTGACATCTTGTGT 540
 QY 541 TCCAGAACATTTGATGGGAAAGAGTTGTAAATGACCAAGATGACTTCCAGAGACTCA 600
 DB 541 TCCAGAACATTTGATGGGAAAGAGTTGTAAATGACCAAGATGACTTCCAGAGACTCA 600
 QY 601 CGCGGAGCTATTAACGAGATATCTCTGTACACACTACACTACCTACAGAGAGAGAG 660
 DB 601 CGCGGAGCTATTAACGAGATATCTCTGTACACACTACACTACCTACAGAGAGAGAG 660
 QY 661 CCACTTTTATTTTCCAAATATCATGCTTACCCAGAGACGCAAGAAATTAACAA 720
 DB 661 CCACTTTTATTTTCCAAATATCATGCTTACCCAGAGACGCAAGAAATTAACAA 720
 QY 721 GGGCAGATTTTACCTTATGAGCAAGGAGGATCAGCTGAGAGATTCACAGCATCCCA 780
 DB 721 GGGCAGATTTTACCTTATGAGCAAGGAGGATCAGCTGAGAGATTCACAGCATCCCA 780

QY 781 CTCAGTCAAAAGCTACCCCAACCATCATCTTTCAGACAGTGGCCCAAAACAGAGACAGCTC 840
 DB 781 CTCAGTCAAAAGCTACCCCAACCATCATCTTTCAGACAGTGGCCCAAAACAGAGACAGCTC 840
 QY 841 CTCAGTATATCCTTTATTCAGATTTCTTGACCGGACCGAGCCGCTCTGCAAAATCCAGGGA 900
 DB 841 CTCAGTATATCCTTTATTCAGATTTCTTGACCGGACCGAGCCGCTCTGCAAAATCCAGGGA 900
 QY 901 GTGGGAGATTAAGCTATATGAGCTTCTCTAGAGCTTCTGTGCGACAGCTCCAACTCCA 960
 DB 901 GTGGGAGATTAAGCTATATGAGCTTCTCTAGAGCTTCTGTGCGACAGCTCCAACTCCA 960
 QY 961 ACTGCATCACCTGGGAGGACCAAAATGGGAGTTCAGATGACAGACCTGATGAAGTGG 1020
 DB 961 ACTGCATCACCTGGGAGGACCAAAATGGGAGTTCAGATGACAGACCTGATGAAGTGG 1020
 QY 1021 CTCGGGCTTGGGAGAGAGAGAAAGCAACTAACGTGAATGATGCAAAATCAGCCGTCG 1080
 DB 1021 CTCGGGCTTGGGAGAGAGAGAAAGCAACTAACGTGAATGATGCAAAATCAGCCGTCG 1080
 QY 1081 CACTTCGCTACTACTATGACAAAATATTTATGACTAAAGTTCATGTAACGCTATGCTCT 1140
 DB 1081 CACTTCGCTACTACTATGACAAAATATTTATGACTAAAGTTCATGTAACGCTATGCTCT 1140
 QY 1141 ACAATTTTATTTCCAGGAAATGCTCAGAGCCCTCCAGCTCACTCCAGATATCCA 1200
 DB 1141 ACAATTTTATTTCCAGGAAATGCTCAGAGCCCTCCAGCTCACTCCAGATATCCA 1200
 QY 1201 TGTACAAATATACCATAGACAGCTCCCTACATGAGTTCCTACCATGACAGCCCGAGAA 1260
 DB 1201 TGTACAAATATACCATAGACAGCTCCCTACATGAGTTCCTACCATGACAGCCCGAGAA 1260
 QY 1261 TGAACCTTTAGCTCCCATCCCTCCCTGCTTCCCGTAACCTATCCAGCTTTTGTGCTG 1320
 DB 1261 TGAACCTTTAGCTCCCATCCCTCCCTGCTTCCCGTAACCTATCCAGCTTTTGTGCTG 1320
 QY 1321 CCCCTATCCATCTGGAATTCACCACTGAGGATCTACCCCAATACAGAGCTGCGAG 1380
 DB 1321 CCCCTATCCATCTGGAATTCACCACTGAGGATCTACCCCAATACAGAGCTGCGAG 1380
 QY 1381 CTGCTCATATGCTTCCCATCTTGGCAGCTACTACTAATGAGGGAAGAAAGAGCGCC 1440
 DB 1381 CTGCTCATATGCTTCCCATCTTGGCAGCTACTACTAATGAGGGAAGAAAGAGCGCC 1440
 QY 1441 AAGAAAA 1447
 DB 1441 AAGAAAA 1447

RESULT 2
 AAX26552
 ID AAX26552 standard; DNA; 1528 BP.
 XX
 AC AAX26552;
 DT 14-JUN-1999 (first entry)
 DE DNA encoding chicken c-ery protein.
 XX
 KW Chicken; C-11 protein; cell calcification inhibiting activity;
 KW cell calcification inhibiting agent; c-ery protein; arthritis deformans;
 OS ossification; spinal column ligament; ss.
 XX
 OS Gallus sp.
 XX
 FH Key Location/Qualifiers
 FT CDS 63..1499
 FT /*tag= a
 XX
 PN JPI1075871-A.
 XX
 PD 23-MAR-1999.

XX		29-JAN-1998;	98JP--0166076.	
PE				
XX		20-JUN-1997;	97US--0050297.	
FR		18-JUN-1997;	97US--0878177.	
XX		(CHUS) CHUGAI PHARM CO LTD.		
PA		(UYFE-) UNIV PENNSYLVANIA.		
XX				
DR		WPI; 1999-257708/22.		
DR		P-PSDB; MAY01521.		
XX				
PT		An active protein for inhibiting cell calcification - useful for		
PT		measuring the calcification of a cell, for diagnosing arthritis		
PT		deformans or ossification of spinal column ligament		
XX				
PS				
XX		Disclosure; Page 8-9; 15pp; Japanese.		
CC				
CC		The present sequence encodes a chicken c-erb protein. The specification		
CC		also describes a chicken C-11 protein (AA01520) which has cell		
CC		calcification inhibiting activity and a cell calcification inhibiting		
CC		agent containing c-erb protein. The proteins are used for measuring the		
CC		calcification of a cell, for diagnosing arthritis deformans or		
CC		ossification of spinal column ligament.		
XX				
SO		Sequence 1528 BP; 464 A; 395 C; 330 G; 339 T; 0 other:		
Query Match		93.7%; Score 1356; DB 20; Length 1528;		
Best Local Similarity		94.7%; Pred. NO. 0;		
Matches 1447;		Conservative 0; Mismatches 0; Indels 81; Gaps		
OY	1	GAAATTCGGCGAAGCAATAATTATTTATTTACGAATTTATGACGATCAATAATCTTGATCACA	60	
Dd	1	GAAATTCGGCGAAGCAATAATTATTTATTTACGAATTTATGACGATCAATAATCTTGATCACA	60	
OY	61	TTATGCGAAGCATCTTATTAAGAAGCATTTATCAGTGTGATGATNAGACCAGTCTTGTG	120	
Dd	61	TTATGCGAAGCATCTTATTAAGAAGCATTTATCAGTGTGATGATNAGACCAGTCTTGTG	120	
OY	121	AGTGTGCCTTAGCGATGCCGCCACCTTGCAAGACAGAAAATGACAGCTTCCTTCACGTG	180	
Dd	121	AGTGTGCCTTAGCGATGCCGCCACCTTGCAAGACAGAAAATGACAGCTTCCTTCACGTG	180	
OY	181	AATATGGCGAACAATCAAAGATGAGCCCGCGGTTCGCCAGCAGCACTGTTATCACAGC	240	
Dd	181	AATATGGCGAACAATCAAAGATGAGCCCGCGGTTCGCCAGCAGCACTGTTATCACAGC	240	
OY	241	CCCCGGCCAGATTTACCAATTAGATGAGTGATACCCAAACGAGTTATGGGTCAAGGA	300	
Dd	241	CCCCGGCCAGATTTACCAATTAGATGAGTGATACCCAAACGAGTTATGGGTCAAGGA	300	
OY	301	ATTTCACCTGATGACTGACAGCCGTGGCAAAAAGAGGAAAAATGTTAGCATTCAGACATG	360	
Dd	301	ATTTCACCTGATGACTGACAGCCGTGGCAAAAAGAGGAAAAATGTTAGCATTCAGACATG	360	
OY	361	TGCGATGATGAATATGGAAGCTACATGGAAGAGACATATTCGCCCTCCAAAATATGACA	420	
Dd	361	TGCGATGATGAATATGGAAGCTACATGGAAGAGACATATTCGCCCTCCAAAATATGACA	420	
OY	421	CCAATGAAGCAAGATTTATGTGCCAGCAGATCCCTACGTTATGAGACACAGACCAAGTAC	480	
Dd	421	CCAATGAAGCAAGATTTATGTGCCAGCAGATCCCTACGTTATGAGACACAGACCAAGTAC	480	
OY	481	GGCAGTGGCTGAGTGGCGCATGGAAGAGATGATGTCCTTCACAGCGGACATCTTGTGTG	540	
Dd	481	GGCAGTGGCTGAGTGGCGCATGGAAGAGATGATGTCCTTCACAGCGGACATCTTGTGTG	540	
OY	541	TCCAGAACATTTGATGGGAAAAGAGTTGTGTAATAATGACCAAAAGATGACTTCCAGAGACTCA	600	
Dd	541	TCCAGAACATTTGATGGGAAAAGAGTTGTGTAATAATGACCAAAAGATGACTTCCAGAGACTCA	600	
OY	601	CCCCGAGCTATTAACGCAAGATATCCCTCTCACACCTTAACATCACTCCAGAGAGA-----	654	
Dd	601	CCCCGAGCTATTAACGCAAGATATCCCTCTCTCACACCTTAACATCACTCCAGAGAGA-----	654	

Dd	601	CGCCGACGTATTAAGCAGATATCTCTCTGCAACACTACACTACACTCAGAGAGACTCTC	660
Qy	655	-----	654
Dd	661	TTTCACATTTTGACTTCAGATGATGTGTAAGGCCCTTACAAAACCTCTCCAGCGTTAATGC	720
Qy	655	-----GAGGAGCCACTTTTATTTTCCAAATACATCAGTTTACCAGAG	699
Dd	721	ATGCTAGAAACACAGGAGGCCACTTTTATTTTTCCAATACATCAGTTTACCAGAGAG	780
Qy	700	CAAGCAGAAAGATTAACAACAAGGCCAGATTACCTTATGAGCAAGCGAGATCAGCGT	759
Dd	781	CAAGCAGAAAGATTAACAACAAGGCCAGATTACCTTATGAGCAAGCGAGATCAGCGT	840
Qy	760	GGACGAGTCACAGCCATCCACTCAGTCAAAAGCTAACCCACATCATCTTCAACAGTGC	819
Dd	841	GGAGAGTCACAGCCATCCACTCAGTCAAAAGCTAACCCACATCATCTTCAACAGTGC	900
Qy	820	CCAAAACAGAAAGCCACCGTCTCAGTTAGTTCCTTATCAGATTCTTGAGCCGACACGA	879
Dd	901	CCAAAACAGAAAGCCACCGTCTCAGTTAGTTCCTTATCAGATTCTTGAGCCGACACGA	960
Qy	880	GCCGCTTGGAAATTCAGGGAGTGGGACAGTACAGTATGAGAGTTCACAGGAGTTC	939
Dd	961	GCCGCTTGGAAATTCAGGGAGTGGGACAGTACAGTATGAGAGTTCACAGGAGTTC	1020
Qy	940	TGTGGCAGCGTCCAACTCCAACTGCATCACCCTGGGAGGGCCAAATGGGGAGTTCAAGA	999
Dd	1021	TGTGGCAGCGTCCAACTCCAACTGCATCACCCTGGGAGGGCCAAATGGGGAGTTCAAGA	1080
Qy	1000	TGACAGAGCCCTGATGAAAGTGGCTTCGGCGTTGGGGAGAGAGAAAGCAAACTTAATCA	1059
Dd	1081	TGACAGAGCCCTGATGAAAGTGGCTTCGGCGTTGGGGAGAGAGAAAGCAAACTTAATCA	1140
Qy	1060	ACTATGACAAATCAGCGCTGACACTTCGCTAGTACTATGACAAAAATATTATGACTAAG	1119
Dd	1141	ACTATGACAAATCAGCGCTGACACTTCGCTAGTACTATGACAAAAATATTATGACTAAG	1200
Qy	1120	TTTCATGCTTAAAGCGTATGCCCTACAAATTTGATTTCCAGGAAATGCTCAGGCCCTCAGC	1179
Dd	1201	TTTCATGCTTAAAGCGTATGCCCTACAAATTTGATTTCCAGGAAATGCTCAGGCCCTCAGC	1260
Qy	1180	CTCAGCTCCGAAATCATTCATGTACAAATACCCATCAACCTCCCTTCATGAGTTCTT	1239
Dd	1261	CTCAGCTCCGAAATCATTCATGTACAAATACCCATCAACCTCCCTTCATGAGTTCTT	1320
Qy	1240	ACCATGCAACCCCCACAAGATGAGCTTGTTAGCTCCCATCCCGCTGGTTGCCCGTAA	1299
Dd	1321	ACCATGCAACCCCCACAAGATGAGCTTGTTAGCTCCCATCCCGCTGGTTGCCCGTAA	1380
Qy	1300	CCCTCATCAGGCTTTTGTGTGCCCTTAATTCATCACTGGAATTCACCAACTGGAGGACTCT	1359
Dd	1381	CCCTCATCAGGCTTTTGTGTGCCCTTAATTCATCACTGGAATTCACCAACTGGAGGACTCT	1440
Qy	1360	ACCCCATATCCAGGCTGCCAGCTGCTCATATGCTTCATCCATTTGGCAGCTACTAATAAG	1419
Dd	1441	ACCCCATATCCAGGCTGCCAGCTGCTCATATGCTTCATCCATTTGGCAGCTACTAATAAG	1500
Qy	1420	TGGGGAAGAAAGAAAGCGGCAAGAAA 1447	
Dd	1501	TGGGGAAGAAAGAAAGCGGCAAGAAA 1528	
RESULT 3			
AA050644			
AA050644 standard; cDNA; 2938 BP.			
AA050644;			
26-MAY-1994 (first entry)			
Human Hum-Fil-1 gene clone BM025.			

KW chromosomal translocation: chimeric; chimeric; Ewing sarcoma;
KW Ews gene; malignant melanoma; hum-Flt-1;
KW primitive peripheral neuroectodermal tumour; human chromosome 11;
KW human chromosome 22; ss.
OS Homo sapiens.
XX
XX
XX Key Location/Qualifiers
XX CDS 143..1501
XX FT /tag= a
XX FT /Product= HUM-FLI-1
XX FT 2908..2913
XX FT /tag= b
XX
XX WO933549-A.
XX
XX 25-NOV-1993.
XX
XX 19-MAY-1993; 93WO-FR00494.
XX
XX 20-MAY-1992; 92FR-0006123.
XX
XX (CNRS) CNRS CENT NAT RECH SCI.
XX
XX Aurias A, Delattre O, Desmaza C, Melot T, Peter M;
XX Plougastel B, Thomas G, Zucman J;
XX
XX WPI: 1993-386580/48.
XX P-PSDB: AAR44556.
XX
XX New nucleic acid of EWS gene and its hybrid(s) - contg. gene
XX PT sequence involved in chromosomal translocation, also derived
XX PT mRNA, probes, fusion proteins etc., for diagnosis and treatment
XX PT of Ewing sarcoma and melanoma
XX
XX PS Disclosure: Fig 7; 123pp; French.
XX
XX CC The probe 11KR1 was used to screen a human marrow cDNA library
XX CC (Clontech cat. # HL1058). The clone BM025 was identified and
XX CC sequenced. It represents the entire coding region together with
XX CC 5'- and 3'-UTRs of the Hum-Flt-1 gene.
XX
XX SQ Sequence 2938 BP; 847 A; 692 C; 675 G; 724 T; 0 other;

Query Match 39.0%; Score 563.8; DB 14; Length 2938;
Best Local Similarity 66.5%; Pred. No. 4.8e-175;
Matches 913; Conservative 0; Mismatches 427; Indels 33; Gaps 6;

QY 63 ATGGCAAGCACTTTAAAGAGCATTTATCAGTGTGAGTGAACACACAGCTCTTTGAG 122
DB 143 ATGGACGGGCTTTAAAGAGGCTCTGCTGTGAGGAGCAGCAGCATCTCTTTGAC 202
QY 123 TGTGCTACGATCG--CCCGACCTTGCMAAGACAGAAATGACAGCCTCTTCAGT 179
DB 203 TCGAGCTACGAGCGGCGGACCCATCTCCCAAGCGGACATGACTGCTGGGAGTCTT 262
QY 180 GAATATGGGCAACATCAAGATGAGCCGCGGCTTCCCAAGCAGAGTGGTTATCAG 239
DB 263 GACTACGGGAGCGCCCAAGATCAACCCCTCCACACAGCAGAGAGTGATCAATCAG 322
QY 240 CCCCCCGGAGATTTACATTAAGATGAGTGTACCCCAACCAAGCTTAATGGGTCAAG 299
DB 323 CC-----AGTGAAGGCAACGTCMAAGCGGAGTATGACACATGAATGATCCAGG 373
QY 300 AATTACCTGATGACTGACAGCTGGCAAAAGAGGGAATGTTAGCAGTACAGCAAT 359
DB 374 GAGTCTCCGCTGACTGACGCTTACCAAAATGACAGCAAGTGTGGCGGAGCGAGTCC 433
QY 360 GTTGGAGTAATGAAAGTACATGAAAGAGAGAGCAT--ATTCCGCTCCAAATATG 416
DB 434 AACCCCATGAATCAACAGACTATATGACGAGAGAAGAAATGGCCCTCTCCCAACATG 493
QY 417 ACAACCAATGAACGAAGATTATTGTGCCAGCAGATCTTACGTTATGAGACAGACCAT 476

DB 494 ACCACCAACGAGAGGAGATCATCTGCCGACAGACCCCACTGTGACACAGGAGCAT 553
QY 477 GTACGGCAGTGGCTGAGTGGGAGTGAAGAGTATGTTTCCAGACCTGACATCTTG 536
DB 554 GTGAGGCAATGGCTGAGTGGGCAATTAAGAGATATAGCTTATGAGATGACACATCC 613
QY 537 TTGTTCCAGACATTTGATGAGGAAAGAGTGTGTAATTAATGACCAAGATGACTTCCAG 596
DB 614 TTTTCCAGAACATGATGAGTGAAGCAAACTGTGTAATTAATGAACAGGAGCTCTCCCGC 673
QY 597 CTCAGCGCGAGCTATACGAGATATCCCTCTGACACCTACACTACACTACAGAGAGA 656
DB 674 GCCACCAACCTCTACACAGGAAAGTGTGTTCACACTCAGTTACTCTCAGGAAAG- 732
QY 657 GGAGCCACTTTTATTTTCCAAATACATCAGTTTACCAGAACCAAGCAAAAGATTAACA 716
DB 733 -----TTCACTGCTGGGCTTATATACACCTCCACAGCAGCAACATCTCAGATTGAGT 787
QY 717 ACAAGGCCAGATTTACCTTATGAGCAAGCAGAGATCAGCGTGAACAGTCAACCCAT 776
DB 788 GTCAAGAAAGACCTCTTATGACTCAGTCAAGAGAGACATGAGGCAATTAACATGAAT 847
QY 777 CC---CACTCAGTCAAAAGCTACCAACCATCATCTTCAACAGTCCCAAAACAGAAAG 833
DB 848 TGTGGCTCAACAAAGTCTCTCCCTTGGAGGGGCAACAAAGATCAATGAATACAGAG 907
QY 834 CAGCGTCTCAGTTAGTCTTATCAGATCTTGAACGACGACGACGCTTGGCAAT 893
DB 908 CAACGGCCCAAGCAGATCCGATACAGATCTTGGCCGACGACGACGCTTACGACCAAC 967
QY 894 CAGGAGATGGGAGATATACGCTATGAGCACTTCTTACAGTGTGTCGACAGCTCC 953
DB 968 CTTGGAAGCGGGAGATCTCAGTGTGCAATCTCTCTGAGGCTGTCTCCGACAGCGCC 1027
QY 954 AACTCAACATGATCACTCAGTGTGAGGAGGAGCAAAATGGGAGTCAAGATGACAGCCGTGAT 1013
DB 1028 AACGCCAGCTGTATACCTTGGAGGAGGAGCAACGGGAGTCAAAATGACGAGCCCGAT 1087
QY 1014 GAATGGCTGGGCTGTTGGGAGAGAGAAAGCAAACTTAACATGAACTATGACAAATCTC 1073
DB 1088 GAGGTGGCGGAGGCTGGGCGAGGCGGAAAGCAAGCCCAACATGAAATTCGAAAGCTG 1147
QY 1074 AGCCGTGCACTTGTCTACTATGATGACAAATAATTAATGACTAAGTTCATGTTAAAGC 1133
DB 1148 AGCGGGCCCTCGCTTATGATGATTAATAAATATGACCAAAAGTGCACGGCAAAAG 1207
QY 1134 TATGCTACAAATTTGATTTCCAGGAAATGCTCAGGCGCTCCAGGCTCAGCTCCAGAA 1193
DB 1208 TATGCTTACAAATTTGATTTCCAGGCAATGCTCCAGGCTCTGACGACCAATCCGACCGAG 1267
QY 1194 TCATCATGTACAAATATCCCATCAGACCTCCCTACATGATGATCTTACATGACACGCC 1253
DB 1268 TCGTCCATGTACAAATATCCCTTGTGACATCTCTACATGCTCTCTTCTTACATGACGCCACAG 1327
QY 1254 CAGAAGATGAATTTGTAGTCTCCCATCTCCCTGCTTTGGCCGTAACCTTATCAGCTTT 1313
DB 1328 CAGAAGTGAATTTGTAGTCTCCCTCCCATCTCTCATGAGTCTCTCTCTCTCTCTCTCTCTCT 1387
QY 1314 TTTGCTGCCCTTAATCTCATCTGGAATTCACCAACTGAGGAGGATTCATCCCATATACCA-- 1371
DB 1388 TTTGAGGCGGATCAACAAATCTGACCTCCCAAGGAGGAGGATTCATCCCATATACCA 1447
QY 1372 -----GGCTGCGACCTGCTATATGCTTCCCATCTTGGCACTACTACTTA 1417
DB 1448 GTCCCGCGCATCTTACACACCCAGTGCCTTTCACACTTAAGCAGCTACTACTTA 1500

RESULT 4
ABK84139
ID ABK84139 standard; cDNA; 2957 BP.
XX
AC ABK84139;

14-AUG-2002 (first entry)
Human cDNA differentially expressed in granulocytic cells #710.
Human; ss: granulocytic cell; DNA chip; bacterial infection; viral infection; parasitic infection; protozoal infection; fungal infection; sterile inflammatory disease; psoriasis; rheumatoid arthritis; glomerulonephritis; asthma; thrombosis; cardiac reperfusion injury; renal reperfusion injury; ARDS; adult respiratory distress syndrome; inflammatory bowel disease; Crohn's disease; ulcerative colitis; periodontal disease; granulocyte activation; chronic inflammation; allergy.
Homo sapiens.
MO200228999-A2.
11-APR-2002.
03-OCT-2001; 2001WO-US30821.
03-OCT-2000; 2000US-237189P.
(GENE-) GENE LOGIC INC.
Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;
WPI: 2002-435328/46.
Detecting granulocyte activation by detecting differential expression of genes associated with granulocyte activation, which serves as diagnostic markers that is useful for monitoring disease states and drug toxicity -
Claim 1; SEQ ID No 710; 114pp; English.
The invention relates to detecting (M1) granulocyte (GC) activation (GCA), by detecting the level of expression of gene(s) (Gs) identified by DNA chip analysis as given in the specification, and comparing the expression level to an expression level in an unactivated GC, where differential expression of Gs is indicative of GCA. Also included are modulating (M2) Gs by contacting GC with an agent that alters the expression of at least one gene in Gs; (2) screening (M3) for an agent capable of modulating GCA or an inflammation (especially chronic) in a tissue, an allergic response in a subject, exposure of a subject to a pathogen or sterile inflammatory disease using the gene expression profile; (3) detecting (M4) an inflammation (especially chronic) in a tissue, an allergic response in a subject, exposure of a subject to a pathogen or sterile inflammatory disease, by detecting the level of expression in a sample of the tissue of gene(s) from Gs, where the level of expression of the gene is indicative of inflammation; (4) treating (M5) an inflammation (especially chronic) or in a tissue, an allergic response in a subject, exposure of a subject to a pathogen or sterile inflammatory disease, by contacting a tissue having inflammation with an agent that modulates the expression of gene(s) from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for modulating Gs. M3 is useful for screening an agent capable of modulating GCA preferably in an inflammation in a tissue; M4 is useful for detecting an inflammation (especially chronic) in a tissue, an allergic response in a subject, exposure of a subject to a pathogen or sterile inflammatory disease (e.g. psoriasis, rheumatoid arthritis, glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal reperfusion injury, ARDS, adult respiratory distress syndrome, periodontal bowel disease, Crohn's disease, ulcerative colitis, parasitic infection, protozoal infection, fungal infection and M5 is useful for treating one of the above conditions. The present sequence represents a gene differentially expressed in granulocytes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at http://wipo.int/pub/Published_pct_sequences.

XX	Sequence	2957 BP; 855 A; 698 C; 680 G; 724 T; 0 other:
Query Match	38.9%; Score 562.2; DB 24; Length 2957;	
Best Local Similarity	66.4%; Pred. No. 1,6e-174;	
Matches	912; Conservative 0; Mismatches 428; Indels 33; Gaps	
QY	63 ATGGCAAGCACTATTAAAGAGCAATTATCATGAGTGGTGAAGACCAAGTCCTGTTTGAG	122
DB	173 ATGGACGGGACTATTAAAGAGGCTCTGTGGTGGTAGGAGACCAAGTCCTCTTTGAC	232
QY	123 TGTGCTACGATG----CCCACTTTGCAAAACAGAAATGACAGCTCTCTTCCACT	179
DB	233 TCAGGTACGGAGGGGAGGCCAATCTCCCAAGGCCAGATGACTCTCGGGGAGTCT	292
QY	180 GAATATGGGCAAAATCATCAAAAGATAGCCCCGGGTTTCCAGAGAGACTGTTATCAACG	239
DB	293 GACTACGGGACGCCCCCAAGATCACCCCTTCCACACAGCAGAGATGATCAATCAG	352
QY	240 CCCCCGGCCAGAGTTACCATTTAAGATGGAGTGTAAACCAACAGAGTTAATGGTCAAG	299
DB	353 CC-----AGTAGAGGTCAACGTCACAGCGGAGATATGACATCATGATGGATTCAGG	403
QY	300 AATTACCTGATGACTGACGCGTGGCAAAAGAGGCAAAATGGTTATGCACTTACAGCAAT	359
DB	404 GAGTCTCGGGGACTCAGCGTTAGCAAAATGACGAAGCTGTGGCGGAGCGCAGTCC	463
QY	360 GTTGGAGTGAACCTATGGAAGCTGATGGAAGAGAACAT---ATTCCGCTCCAAATATG	416
DB	464 AACCCTAAGACTCAACAGCTTATATGAGACGAAAGAAATGGCCCCCTCTCCCAACATG	523
QY	417 ACAACCAATGAAACAGAGATTATTTGTGCCAGAGATCTACGTTATGAGACACACCAT	476
DB	524 ACACCAACAGAGAGAGATGATCTGCCGAGACCCACACTGTGAGACACAGAGCAT	583
QY	477 GTACGGCAGTGGCTGGAGTGGGACGTGAAGAGATATGTTCTTCACAGCTGGACATCTTG	536
DB	584 GTGAGGCAATGGCTGGAGTGGGCCATTAAGAGATACAGCTTGATGAGATGACACATTC	643
QY	537 TTTGTTCCAGAACATTGATGGGAAAGATGTGTAAATATGACCAAGATATACCTCCAGAG	596
DB	644 TTTTTCAGAACATGATGGATGGCAAGGAATCTGTAAATGAAACAAAGAGAGATCTCTCCG	703
QY	597 CTCACGGCGACCTTAACGAGATATCCCTGTCCACACTACACTACCTCAGAGAGAG	656
DB	704 GCCACACCTCTTCAACACGGAAGTCTGTTGTCAACCTCATGTTACTTCAGGAAAG--	762
QY	657 GGAGCACTTTATTTTCCAAATACATCACTAGTTTACCAGAAAGCAACGCAAGAAATACA	716
DB	763 -----TTACGTGCGTGGGCTTAATATACAACTCCACACGACCAATCTCAGATTGAGT	817
QY	717 ACAAGGCCAGATTACTTATGACCAAGGAGAGATCAACGCTGGACGATCAGACCAT	776
DB	818 GTCAAGAGAGCCCTTCTTATGACTAGTCAGAGAGGAGCTTGGGCAATTAACATGTAAT	877
QY	777 CC-----CACTGCTCAAAAGCATCCCAACCATCTTCTTCAACAGTGGCCCAAAACAGAAAG	833
DB	878 TCTGGCTCAACAAAGTCTCTCCCTTGGAGGGGCAACAAACATCATGTAAGATACAGAG	937
QY	834 CAGGCTCTCAGTTAGATCTTATACAGATTCTTGGACGACAGCAGCCGTTTCCAAAT	893
DB	938 CAAGGCGCCAGGCCAGATTCGATACAGATCTCTGGGCCACAGCAGATGCTTACCAAC	997
QY	894 CCAGGAGTGGGAGATACAGCTATGGCAGTTCTTACTGAGAGTCTGTGGACAGTCC	953
DB	998 CCTGAAGCGGGAGATCCAGCTGTGGCAATCTCTTGAGACTGCTCTCCAGACGCC	1057
QY	954 AACTCCAACTGCACTACCTTGGAGGCGCAAAATGGGGAATTCATAAGTGCACACCTCAT	1013
DB	1058 AAGCGCAGCTGTATCACCTGGAGGGGACCAACGGGGAATTCATAAATGACGACCCGAT	1117
QY	1014 GAACTGGCTGGCTTTGGGAGAGAGAAAGCAACCTAACATGAATCATACAACTC	1073


```
Db 343 CCGGGCCCTCCGTTACTACTATGACAGACATCATGACCAAGGTCATGGGAAGCGCTA 284
QY 1136 TGCCTACAAATTTGATTTCCAGGAAATCGTCAGGCCCCCTCCAGGCTCCAGAAATC 1195
Db 283 CGCCTACAAAGTTGACCTTCCAGGGATGGCCAGGCCCTCCAGGCCACCCCGGAGTTC 224
QY 1196 ATCATGTACAAATACCATATGACCTCCCTACATGATGTTCTTACATGACACACCCCA 1255
Db 223 ATCTCTGACAAAGTACCCCTGAGACCTCCCTACATGAGGCTCTATACGCGCCACCCACA 164
QY 1256 GAAGATGAACCTTGATGCTCCCATCCCTGCTTTGGCGGTACCTCATGCAAGCTTTT 1315
Db 163 GAAGTGAACCTTTGGGCGCCACCCCTCCAGCCCTCCCTGACATCTTCCAGTTT 104
QY 1316 TGCCTACAAATTTGATTTCCAGGAAATCGTCAGGCCCCCTCCAGGCTCCAGAAATC 1375
Db 103 TGCCTACAAATTTGATTTCCAGGAAATCGTCAGGCCCCCTCCAGGCTCCAGAAATC 1375
QY 1376 GCCAGCTGCTCATATGCTTCCCATCTTGGCAGCTTACTACTAA 1418
Db 43 CCCCACACGACATATGCTTCTCATCTGCGCATTACTACTAA 1
```

RESULT 9

AAK14426/c

ID AAK14426 standard; DNA; 567 BP.

AC AAK14426;

DT 05-NOV-2001 (first entry)

DE Human brain expressed single exon probe SEQ ID NO: 14417.

KW Human; brain expressed exon; gene expression analysis; probe;

KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;

KW epilepsy; cancer; ss.

OS Homo sapiens.

PN W0200157275-A2.

PD 09-AUG-2001.

PE 30-JAN-2001; 2001WO-US00667.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

PA (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483446/52.

XX Single exon nucleic acid probes for analyzing gene expression in human

XX brains -

XX Example 4; SEQ ID NO: 14417; 650bp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid

XX probes which are derived from genomic sequences expressed in the human

XX brain. They can be used to measure gene expression in brain cell samples,

XX which may enable the diagnosis and improved treatment of nervous system

XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,

XX epilepsy and cancers. The present sequence is one of the probes of the

SQ Sequence 567 BP; 114 A; 124 C; 200 G; 129 T; 0 other;

Query Match 26.7%; Score 387; DB 22; Length 567;

Best Local Similarity 83.7%; Pred. No. 5,2e-117;

Matches 438; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

```
QY 896 AGGAGTGGGAGATACGATATGCACTTCCCTACTGAGGCTTGTGTGACAGCTCCAA 955
Db 523 AGGAGTGGGAGATACGATATGCACTTCCCTACTGAGGCTTGTGTGACAGCTCCAA 464
QY 956 CTCCAGCTGCAATCAGCTTGGGAGGAGCAAAATGGGAGTTCAAGATGACAGCCCGAAGA 1015
Db 463 CTCCAGCTGCAATCAGCTTGGGAGGAGCAAAATGGGAGTTCAAGATGACAGCTCCAA 404
QY 1016 AGTGGCTGCGGCTTGGGAGGAGAGAGAAAGCAAACTTAATGATGATGACAAACTCAG 1075
Db 403 GGTGGCCCGGGGCTTGGGAGGAGAGAGAGAAAGCAAACTTAATGATGATGACAAACTCAG 344
QY 1076 CCGTGCACCTTGGCTACTACTATGACAAATAATTTATGACTAAAGTTCAAGTAAAGCTA 1135
Db 343 CCGGCCCCCTCCGTTACTACTATGACAAAGATGACAAAGGTCATGGGAAAGCGCTA 284
QY 1136 TGCCTACAAATTTGATTTCCAGGAAATCGTCAGGCCCCCTCCAGGCTCCAGAAATC 1195
Db 283 CGCCTACAAATTTGATTTCCAGGAAATCGTCAGGCCCCCTCCAGGCTCCAGAAATC 1195
QY 1196 ATCATGTACAAATACCATATGACCTCCCTACATGATGTTCTTACATGACACACCCCA 1255
Db 223 ATCTCTGACAAAGTACCCCTGAGACCTCCCTACATGAGGCTCTATACGCGCCACCCACA 164
QY 1256 GAAGATGAACCTTGATGCTCCCATCCCTGCTTTGGCGGTACCTCATGCAAGCTTTT 1315
Db 163 GAAGTGAACCTTTGGGCGCCACCCCTCCAGCCCTCCCTGACATCTTCCAGTTT 104
QY 1316 TGCCTACAAATTTGATTTCCAGGAAATCGTCAGGCCCCCTCCAGGCTCCAGAAATC 1375
Db 103 TGCCTACAAATTTGATTTCCAGGAAATCGTCAGGCCCCCTCCAGGCTCCAGAAATC 1375
QY 1376 GCCAGCTGCTCATATGCTTCCCATCTTGGCAGCTTACTACTAA 1418
Db 43 CCCCACACGACATATGCTTCTCATCTGCGCATTACTACTAA 1
```

RESULT 10

AAK40160/c

ID AAK40160 standard; DNA; 567 BP.

AC AAK40160;

DT 06-NOV-2001 (first entry)

DE Human bone marrow expressed single exon probe SEQ ID NO: 14717.

KW Human; bone marrow expressed exon; gene expression analysis; probe;

KW microarray; cancer; leukemia; lymphoma; myeloma; ss.

OS Homo sapiens.

PN W0200157276-A2.

PD 09-AUG-2001.

PE 30-JAN-2001; 2001WO-US00668.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI WPI; 2001-488900/53.
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human bone marrow -
PT Example 4; SEQ ID NO: 14717; 658bp + Sequence Listing; English.
PS
XX The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukemia and myeloma. The present sequence is one of
CC the probes of the invention.
XX
SQ Sequence 567 BP; 114 A; 124 C; 200 G; 129 T; 0 other;
Query Match 26.7%; Score 387; DB 22; Length 567;
Best Local Similarity 83.7%; Pred. No. 5.2e-117;
Matches 438; Conservative 0; Mismatches 85; Indels 0; Gaps 0;
QY 896 AGGAGTGGGAGATACAGTATGAGCTTCTACTGAGAGTCTCTGCGAGACGCTCCAA 955
DB 523 AGGAGTGGGAGATACAGTATGAGCTTCTACTGAGAGTCTCTGCGAGACGCTCCAA 464
QY 956 CTCGCAACTGATCAGCTGGAGGAGCAACAATGGAGTTCAGATGACAGACCTGATGA 1015
DB 463 CTCGCAACTGATCAGCTGGAGGAGCAACAATGGAGTTCAGATGACAGACCTGATGA 404
QY 1016 AGTGGCTGGCGCTGGGAGAGAGAAACCAACTACATGATGATGACAAACTCAG 1075
DB 403 GGTGGCCCGGCGCTGGGAGAGAGAAACCAACTACATGATGATGACAAACTCAG 344
QY 1076 CGGTGACCTCGCTACTACTGATGACAAATATATGATGATGATGATGATGATGAT 1135
DB 343 CGGTGACCTCGCTACTACTGATGACAAATATATGATGATGATGATGATGATGATGAT 284
QY 1136 TGGCTACAAATTTGATTTCCAGCGAATCGCTCAGGCGCTCCAGCCCTCCAGATC 1195
DB 283 CGGTGACCTCGCTACTACTGATGACAAATATATGATGATGATGATGATGATGATGAT 224
QY 1196 ATTCATGATACAAATACCATCAGACCTCCCTCATGATGATGATGATGATGATGATGAT 1255
DB 223 ATTCATGATACAAATACCATCAGACCTCCCTCATGATGATGATGATGATGATGATGAT 164
QY 1256 GAGATGAACTTTGTAGTCCGATCCCGCTGCTTGGCCGATGATGATGATGATGATGATGAT 1315
DB 163 GAGATGAACTTTGTAGTCCGATCCCGCTGCTTGGCCGATGATGATGATGATGATGATGAT 104
QY 1316 TGGTGGCCCTTAATCTCATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1375
DB 103 TGGTGGCCCTTAATCTCATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 44
QY 1376 GCCAGCTGCTCATATGCTTCCCATCTTGGCAGCTTACTACTAA 1418
DB 43 CCCGACGAGCATATGCTTCTCATCTGAGGCACTTACTACTAA 1
RESULT 11
AAI20932/c
ID AAI20932 standard; DNA; 567 BP.
XX
XX AAI20932;
AC
XX
XX 12-OCT-2001 (first entry)
DT
XX
XX Probe #10865 for gene expression analysis in human cervical cell sample.
DE
XX
XX Probe: human; microarray; gene expression; cervical epithelial cell;
KW
XX
XX Cervical cancer; 89.

OS Homo sapiens.
XX
XX WO200157278-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00670.
XX
XX 04-FEB-2000; 2000US-0180312.
XX
XX 26-MAY-2000; 2000US-0207456.
XX
XX 30-JUN-2000; 2000US-0608408.
XX
XX 03-AUG-2000; 2000US-0632366.
XX
XX 21-SEP-2000; 2000US-0234687.
XX
XX 27-SEP-2000; 2000US-0236359.
XX
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PI Penn SG, Hanzel DK, Chen W, Rank DR;
DR WPI; 2001-488901/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human cervical epithelial cells -
PT Claim 25; SEQ ID NO 10865; 487bp; English.
PS
XX
XX
XX The present invention relates to human single exon nucleic acid probes
CC (SENPs). The present sequence is one such probe. The SENPs are derived
CC from human HeLa cells. The SENPs can be used to produce a single exon
CC microarray, which can be used for measuring human gene expression in a
CC sample derived from human cervical epithelial cells. By measuring gene
CC expression, the probes are therefore useful in grading and/or staging
CC of diseases of the cervix, notably cervical cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at http://wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 567 BP; 114 A; 124 C; 200 G; 129 T; 0 other;
Query Match 26.7%; Score 387; DB 22; Length 567;
Best Local Similarity 83.7%; Pred. No. 5.2e-117;
Matches 438; Conservative 0; Mismatches 85; Indels 0; Gaps 0;
QY 896 AGGAGTGGGAGATACAGTATGAGCTTCTACTGAGAGTCTCTGCGAGACGCTCCAA 955
DB 523 AGGAGTGGGAGATACAGTATGAGCTTCTACTGAGAGTCTCTGCGAGACGCTCCAA 464
QY 956 CTCGCAACTGATCAGCTGGAGGAGCAACAATGGAGTTCAGATGACAGACCTGATGA 1015
DB 463 CTCGCAACTGATCAGCTGGAGGAGCAACAATGGAGTTCAGATGACAGACCTGATGA 404
QY 1016 AGTGGCTGGCGCTGGGAGAGAGAAACCAACTACATGATGATGATGATGATGATGATGAT 1075
DB 403 GGTGGCCCGGCGCTGGGAGAGAGAAACCAACTACATGATGATGATGATGATGATGATGAT 344
QY 1076 CGGTGACCTCGCTACTACTGATGACAAATATATGATGATGATGATGATGATGATGAT 1135
DB 343 CGGTGACCTCGCTACTACTGATGACAAATATATGATGATGATGATGATGATGATGAT 284
QY 1136 TGGCTACAAATTTGATTTCCAGCGAATCGCTCAGGCGCTCCAGCCCTCCAGATC 1195
DB 283 CGGTGACCTCGCTACTACTGATGACAAATATATGATGATGATGATGATGATGATGAT 224
QY 1196 ATTCATGATACAAATACCATCAGACCTCCCTCATGATGATGATGATGATGATGATGAT 1255
DB 223 ATTCATGATACAAATACCATCAGACCTCCCTCATGATGATGATGATGATGATGATGAT 164
QY 1256 GAGATGAACTTTGTAGTCCGATCCCGCTGCTTGGCCGATGATGATGATGATGATGATGAT 1315
DB 163 GAGATGAACTTTGTAGTCCGATCCCGCTGCTTGGCCGATGATGATGATGATGATGATGAT 104
QY 1316 TGGTGGCCCTTAATCTCATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1375

DB	103	TCCTGCCCCCAACCCACTACTGSAATTCACCAACTGGGGTATATATACCCCAACACTAGGCT	44
OY	1376	GCCAGCTGCTCATATATGCTTCCCATCTTGGACACTACTACTAA	1418
DB	43	CCCCACAGCCATATGCTTCTCATCTGCGCACTACTACTAA	1
RESULT 12			
ID	AA146176/C		
XX	AA146176 standard; DNA; 567 BP.		
AC	AA146176;		
DX	17-OCT-2001 (first entry)		
DE	Probe #14662 used to measure gene expression in human placenta sample.		
XX	Probe; microarray; human; placenta; antenatal diagnosis;		
KM	genetic disorder; ss.		
OS	Homo sapiens.		
PN	W0200157272-A2.		
XX	09-AUG-2001.		
XX	30-JAN-2001; 2001WO-US000663.		
XX	04-FEB-2000; 2000US-0180312.		
PR	26-MAY-2000; 2000US-0207456.		
PR	30-JUN-2000; 2000US-0608408.		
PR	03-AUG-2000; 2000US-0632366.		
PR	21-SEP-2000; 2000US-0234687.		
PR	27-SEP-2000; 2000US-0236359.		
PR	04-OCT-2000; 2000GB-0024263.		
XX	(MOLE-) MOLECULAR DYNAMICS INC.		
XX	Penn SG, Hanzel DK, Chen W, Rank DR;		
XX	WPI: 2001-488897/53.		
DR	Human genome-derived single exon nucleic acid probes useful for		
XX	analyzing gene expression in human placenta -		
PT	Claim 25; SEQ ID NO 14862; 654bp; English.		
XX	The present invention relates to single exon nucleic acid probes (SENP).		
CC	The present sequence is one such probe. The probes are useful for		
CC	producing a microarray for predicting, measuring and displaying gene		
CC	expression in samples derived from human placenta. The probes are useful		
CC	for antenatal diagnosis of human genetic disorders.		
XX	Sequence 567 BP; 114 A; 124 C; 200 G; 129 T; 0 other;		
XX	Query Match	26.7%; Score 387; DB 22; Length 567;	
XX	Best Local Similarity	83.7%; Pred. No. 5,2e-117;	
XX	Matches 438; Conservative 0; Mismatches 85; Indels 0; Gaps 0;		
OY	896	AGGAGTGGGGGAGATACACTATGCGAGTCTCTACTGAGCTTGTGCGAGAGCTCCAA	955
DB	523	AGGCGAGTGGCCAGATCCAGTCTTGGCAGTCTCTCGAGCTCTGTCGGACAGCTCCAA	464
OY	956	CTCCACATCGCATCACTCTGGGAGGCAACAATGGGAGTTCAGATGACAGACCTGATGA	1015
DB	463	CTCCAGCTGCAATCACCCTGGGAAGGACCAACGGGAGTTCAGATGACGATCCGACGA	404
OY	1016	AGTGCCTGGCGCTTGGGGAGAGAGGAAAGCAAACTTACATGACTATGACAAACTCG	1075
DB	403	GGTGGCCCGGCGCTGGGGAGAGCGGAGAGCAAAACCAACATGACTATGATTAAGTCTAG	344
OY	1076	CGTGCACTTGGCTACTACTATGACAAAAATATATGACTTAAAGTTCAATGTTAAAGCTTA	1135

Db	343	CCGGCCCTCCGTTACTACTATGACAAAGACATCATGACCAAGTCCATGGGAAAGCGTA	284
Qy	1136	TGCCCCAATATTTATTTCCACGGAATCGTCCAGGCCCTCCAGCCCTCCAGAAATC	1195
Db	283	CGCCTACAAAGTTGACACTTCCACGGGAATCGCCAGGCCCTCCAGCCCAACCCCGGAACTC	224
Qy	1136	ATCCATGTACAAATACCCATCAGACCTCCCTTCATGAGTTCCCTACATGACACACCCCA	1255
Db	223	ATCTCTGTACAAAGTAAACCCCTCAGACCTCCCGTCAATAGGGCTCTCATCAGCCCAACA	164
Qy	1236	GAAGTGAACCTTTACTCTCCCATCCCTCGTGTGGCCGTAACCTCATCCAGTTTTT	1315
Db	163	GAAGTGAACCTTTGTGGCGCCCAACCTCCAGCCCTCCCGTACATCTTCCAGTTTTT	104
Qy	1316	TGCTGCCCTTATCTCATCTGGAATTCACCACTGGAGGACATCTACCCCAATACCAAGCT	1375
Db	103	TGCTGCCCAACCAACCATACCTGGAATTCACCACTGGGGGTAATATACCCCAACACTAAGCT	44
Qy	1376	GCCAGCTGCTCATATGCTTCCCATCTTGGACACTACTATA	1418
Db	43	CCCCACGACCATATGCTTCTCTCATCTGGGCACTTACTACTAA	1
RESULT 13			
ID	AA106643/c		
XX	AA106643	standard; DNA; 567 BP.	
XX	AA106643;		
AC	AA106643;		
DT	09-OCT-2001	(first entry)	
XX			
DE	Probe #6634 used to measure gene expression in human breast sample.		
XX			
KW	Probe: human; breast disease; breast cancer; development disorder; ss;		
XX	inflammatory disease; proliferative breast disease; non-carcinoma tumour		
OS	Homo sapiens.		
XX			
PN	WO200157270-A2.		
XX			
PD	09-AUG-2001.		
XX			
PF	29-JAN-2001; 2001WO-US00661.		
XX			
PR	04-FEB-2000; 2000US-0180312.		
PR	26-MAY-2000; 2000US-0207456.		
PR	30-JUN-2000; 2000US-0608408.		
PR	03-AUG-2000; 2000US-0632366.		
PR	21-SEP-2000; 2000US-0234687.		
PR	27-SEP-2000; 2000US-0236359.		
PR	04-OCT-2000; 2000GB-0024263.		
XX			
PA	(MOLE-) MOLECULAR DYNAMICS INC.		
XX			
PI	Penn SG, Hanzel DK, Chen W, Rank DR;		
XX			
DR	WPI: 2001-476286/51.		
XX			
PT	Novel single exon nucleic acid probe used to measuring gene expression		
XX	in a human breast -		
PS	Claim 25; SEQ ID NO 6634; 322pp; English.		
XX			
CC	The present invention relates to novel single exon nucleic acid probes,		
CC	The present sequence is one such probe. The probes are useful for		
CC	measuring human gene expression in a human breast sample, where the probe		
CC	hybridises at high stringency to a nucleic acid expressed in the human		
CC	breast. The probes are useful for predicting, diagnosing, grading,		
CC	staging, monitoring and prognosing diseases of the human breast,		
CC	particularly those diseases with polygenic aetiology. The diseases		
CC	include: breast cancer, disorders of development, inflammatory diseases		
CC	of the breast, fibrocystic changes, proliferative breast disease and		

CC non-carcinoma tumours.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX

Seq Sequence 567 BP; 114 A; 124 C; 200 G; 129 T; 0 other;

Query Match 26.7%; Score 387; DB 22; Length 567;
 Best Local Similarity 83.7%; Pred. No. 5.2e-117;
 Matches 438; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 896 AGGAGTGGGCGAGATACATGACATGCTTCTGAGAGCTTCTGCGAGACGCTCCAA 955
 DB 523 AGGAGTGGGCGAGATACATGACATGCTTCTGAGAGCTTCTGCGAGACGCTCCAA 464
 QY 956 CTCACACTGATCAGCTGGGAGGAGCAAAATGGGAGTTCAAGATGACAGCCCTGATTA 1015
 DB 463 CTCACAGCTGATCAGCTGGGAGGAGCAAAATGGGAGTTCAAGATGACAGCCCTGATTA 404
 QY 1016 AGTGGCTGGGCTGGGAGGAGCAAAATGGGAGTTCAAGATGACAGCCCTGATTA 1075
 DB 403 GGTGGCCCGGCTGGGAGGAGCAAAATGGGAGTTCAAGATGACAGCCCTGATTA 344
 QY 1076 CGTGGCTGGGCTGGGAGGAGCAAAATGGGAGTTCAAGATGACAGCCCTGATTA 1135
 DB 343 CGGCGCCCTGGGCTGGGAGGAGCAAAATGGGAGTTCAAGATGACAGCCCTGATTA 284
 QY 1136 TGCCTACAAATTTGATTTTCACAGGATGCTGAGGCTCCAGACCTCAGCTCCAGATTC 1195
 DB 283 GCGCTACAAATTTGATTTTCACAGGATGCTGAGGCTCCAGACCTCAGCTCCAGATTC 224
 QY 1196 ATCCATGTAACAATATCCATGACAGCTCCCTCAGATGATGCTCCATGACAGCCCA 1255
 DB 223 ATCTGTGTCAAGTATCCCTCAGACCTCCCTGATGATGCTCCATGACAGCCCA 164
 QY 1256 GAAGATGACTTGTGATGCTCCCTCAGACCTCCCTGATGATGCTCCATGACAGCCCA 1315
 DB 163 GAAGATGACTTGTGATGCTCCCTCAGACCTCCCTGATGATGCTCCATGACAGCCCA 104
 QY 1316 TGCCTACAAATTTGATTTTCACAGGATGCTGAGGCTCCAGACCTCAGCTCCAGATTC 1375
 DB 103 TGCCTACAAATTTGATTTTCACAGGATGCTGAGGCTCCAGACCTCAGCTCCAGATTC 44
 QY 1376 GCGAGCTGCTCATATGCTTCCATGCTTGGGACCTACTACTAA 1418
 DB 43 CCCACACAGCCATATGCTTCTCATCTGCGCACTACTACTAA 1

RESULT 14

ID ABS14215/C
 ABS14215 standard; DNA; 567 BP.

XX ABS14215;
 DE 19-AUG-2002 (first entry)

XX Human genome-derived single exon probe ORF from lung SEQ ID No 14206.

XX Human; ds; single exon probe; asthma; lung cancer; COPD; ILD;
 KW Chronic obstructive pulmonary disease; interstitial lung disease;
 KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
 KW tuberculous scleriosis; Gaucher's disease; Niemann-Pick disease;
 KW Heremansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
 KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karsenger syndrome;
 KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
 KW primary ciliary dyskinesia; pulmonary hypertension;
 KW hyaline membrane disease; open reading frame; ORF.

XX Homo sapiens.

XX MO200186003-A2.

XX 15-NOV-2001.

XX 30-JAN-2001; 2001WO-0500665.
 XX
 XX 04-FEB-2000; 2000US-180312P.
 PR 26-MAY-2000; 2000US-207456P.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-234687P.
 PR 27-SEP-2000; 2000US-236359P.
 PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2002-114183/75.

PT Spatially-addressable set of single exon nucleic acid probes, used to
 PT measure gene expression in human lung samples -

XX Claim 4; SEQ ID No 14206; 634bp; English.

CC The invention relates to a spatially-addressable set of single exon
 CC nucleic acid probes for measuring gene expression in a sample derived
 CC from human lung comprising single exon nucleic acid probes having one of
 CC 12614 nucleic acid sequences mentioned in the specification, or their
 CC complements or the 12387 open reading frames derived from the 12614
 CC probes. Also included are a microarray comprising the novel set of
 CC probes; the novel set of probes which hybridise at high stringency to a
 CC nucleic acid expressed in the human lung; measuring gene expression in a
 CC sample derived from human lung, comprising (a) contacting the array with
 CC a collection of detectably labeled nucleic acids derived from human lung
 CC mRNA, and (b) measuring the label detectably bound to each probe of
 CC the array; identifying exons in a eukaryotic genome, comprising
 CC (a) algorithmically predicting at least one exon from genomic sequences
 CC of the eukaryote; and (b) detecting specific hybridisation of detectably
 CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
 CC having a fragment identical to the predicted exon, the probe is included
 CC in the above mentioned microarray; assigning exons to a single gene,
 CC comprising (a) identifying exons from genomic sequence by the method
 CC above and (b) measuring the expression of each of the exons in several
 CC tissues and/or cell types using hybridisation to a single exon
 CC microarray having a probe with the exon, where a common pattern of
 CC expression of the exons in the tissues and/or cell types indicates that
 CC the exons should be assigned to a single gene; a peptide comprising one
 CC of 12011 sequences, mentioned in the specification, or encoded by the
 CC probes/open reading frames (ORF). The probes are used for gene
 CC expression analysis, and for identifying exons in a gene, particularly
 CC using human lung derived mRNA and for the study of lung diseases
 CC such as asthma, lung cancer, chronic obstructive pulmonary disease
 CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary
 CC fibrosis, neurofibromatosis, tuberculous scleriosis, Gaucher's disease,
 CC Niemann-Pick disease, Heremansky-Pudlak syndrome, sarcoidosis, pulmonary
 CC haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis,
 CC pulmonary alveolar proteinosis, Karsenger syndrome, fibrocystic
 CC pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension
 CC and hyaline membrane disease. The present sequence is a single exon
 CC probe open reading frame of the invention.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

Seq Sequence 567 BP; 114 A; 124 C; 200 G; 129 T; 0 other;

Query Match 26.7%; Score 387; DB 24; Length 567;
 Best Local Similarity 83.7%; Pred. No. 5.2e-117;
 Matches 438; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 896 AGGAGTGGGCGAGATACATGACATGCTTCTGAGAGCTTCTGCGAGACGCTCCAA 955
 DB 523 AGGAGTGGGCGAGATACATGACATGCTTCTGAGAGCTTCTGCGAGACGCTCCAA 464

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 9, 2002, 10:20:16 ; Search time 95.8082 Seconds
(without alignments)
4631.768 Million cell updates/sec

Title: US-09-902-772-1

Perfect score: 1447
1 gaattccgcgaacgaataat.....gaaagaagcgcaagaanaa 1447

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: Issued_Patents_NA:*
2: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*
6: /cgn2_6/ptodata/1/ina/PCTUS.COMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1447	100.0	1447	4 US-08-878-177-1	Sequence 1, Appli
2	1356	93.7	1528	4 US-08-878-177-3	Sequence 3, Appli
3	563.8	39.0	2938	2 US-08-343-443B-3	Sequence 3, Appli
4	180	12.4	1752	4 US-09-360-779-1	Sequence 1, Appli
5	180	12.4	1752	4 US-09-435-335-1	Sequence 1, Appli
6	124.6	8.6	1604	1 US-08-306-691B-43	Sequence 43, Appli
7	124.6	8.6	1604	5 PCT-US93-06251-9	Sequence 9, Appli
8	122.2	8.4	2268	3 US-09-344-579-1	Sequence 1, Appli
9	117.4	8.1	2667	2 US-08-459-412A-1	Sequence 1, Appli
10	117.4	8.1	2667	2 US-09-021-715-1	Sequence 1, Appli
11	115.2	8.0	328	2 US-08-343-443B-5	Sequence 5, Appli
12	96	6.6	2064	3 US-08-875-944B-1	Sequence 1, Appli
13	96	6.6	2064	4 US-09-116-049-3	Sequence 3, Appli
14	95.2	6.6	2410	2 US-08-780-835B-1	Sequence 1, Appli
15	95.2	6.6	2410	4 US-09-303-268-1	Sequence 1, Appli
16	95.2	6.6	2410	4 US-09-116-049-1	Sequence 1, Appli
17	88.6	6.1	2266	2 US-09-213-767-1	Sequence 1, Appli
18	87.4	6.0	2544	2 US-08-469-412A-6	Sequence 6, Appli
19	87.4	6.0	2544	4 US-09-021-715-6	Sequence 5, Appli
20	67.6	4.7	65042	4 US-09-784-316-3	Sequence 3, Appli
21	59.8	4.1	852	4 US-09-020-956-44	Sequence 44, Appli
22	59.8	4.1	852	4 US-09-030-607-44	Sequence 44, Appli
23	59.8	4.1	852	4 US-09-605-785-44	Sequence 44, Appli
24	59.8	4.1	852	4 US-09-439-313-44	Sequence 44, Appli
25	59.8	4.1	852	4 US-09-352-616A-44	Sequence 44, Appli
26	59.8	4.1	852	4 US-09-232-149A-44	Sequence 44, Appli
27	58.2	4.0	5427	3 US-09-009-913-2	Sequence 2, Appli

28	58.2	4.0	5510	3 US-09-009-913-3	Sequence 3, Appli
29	58.2	4.0	5667	3 US-09-009-913-4	Sequence 4, Appli
30	55.2	3.8	2975	1 US-08-368-281-1	Sequence 1, Appli
31	55.2	3.8	3240	1 US-08-368-281-3	Sequence 1, Appli
32	53.6	3.7	1920	1 US-08-746-789A-1	Sequence 1, Appli
33	53.4	3.7	1905	4 US-09-055-113-2	Sequence 2, Appli
34	48.4	3.3	7218	1 US-08-232-463-14	Sequence 14, Appli
35	47.6	3.3	848	3 US-09-009-913-338	Sequence 338, App
36	47.4	3.3	7218	1 US-08-232-463-14	Sequence 8, Appli
37	38.8	2.7	2280	3 US-09-009-913-8	Sequence 10, Appli
38	38.8	2.7	2428	3 US-09-009-913-6	Sequence 3, Appli
39	38.8	2.7	2498	2 US-08-469-412A-3	Sequence 10, Appli
40	35.2	2.4	2432	2 US-08-469-412A-3	Sequence 3, Appli
41	35.2	2.4	2432	4 US-09-021-715-3	Sequence 1, Appli
42	33.4	2.3	1212	3 US-09-046-578-1	Sequence 1, Appli
43	32.6	2.3	1447	4 US-08-878-177-1	Sequence 1, Appli
44	32.6	2.3	1528	4 US-08-878-177-3	Sequence 3, Appli
45	32.6	2.3	12886	4 US-09-453-702B-14	Sequence 14, Appli

ALIGNMENTS

```
RESULT 1
US-08-878-177-1
; Sequence 1, Application US/08878177
; Patent No. 6294354
; GENERAL INFORMATION:
; APPLICANT: Chugai Selyaku Kabushiki Kaisha, 5001, Iwamoto et
; TITLE OF INVENTION: Cell Calcification Suppressing Proteins and Genes of
; FILE REFERENCE: chugai selyaku kabushiki kaisha 5001
; CURRENT APPLICATION NUMBER: US/08/878,177
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1447
; TYPE: DNA
; ORGANISM: C-11 gene, c-ery gene w/ deletion, chicken DNA
US-08-878-177-1

Query Match      100.0%; Score 1447; DB 4; Length 1447;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1447; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAATCCGCGAAGCAATATTTATTTAGCAATTTTACGATCAATATCTGTATCACA 60
DB 1 GAATTCGCGAAGCAATATTTATTTAGCAATTTTACGATCAATATCTGTATCACA 60
QY 61 TTATGCGAAGCACTTTTAAAGAAAGCATTTATCATGAGTGAAGAAAGCAAGCTCTGTTG 120
DB 61 TTATGCGAAGCACTTTTAAAGAAAGCATTTATCATGAGTGAAGAAAGCAAGCTCTGTTG 120
QY 121 AGTGTGCTACGATCGCCGCCACCTTGCAAAAGCAAAATGACAGCTCTCTTCAGTG 180
DB 121 AGTGTGCTACGATCGCCGCCACCTTGCAAAAGCAAAATGACAGCTCTCTTCAGTG 180
QY 121 AGTGTGCTACGATCGCCGCCACCTTGCAAAAGCAAAATGACAGCTCTCTTCAGTG 180
DB 121 AGTGTGCTACGATCGCCGCCACCTTGCAAAAGCAAAATGACAGCTCTCTTCAGTG 180
QY 181 AATATGCGCAAAATCAAAAGATGAGCCCGCGGTTCCCGAGAGACATGTTATACAGC 240
DB 181 AATATGCGCAAAATCAAAAGATGAGCCCGCGGTTCCCGAGAGACATGTTATACAGC 240
QY 241 CCCCCGCGAAGTTACATTAAGATGAGTGTACCCAAACCAAGTTAATGGGTCAAGGA 300
DB 241 CCCCCGCGAAGTTACATTAAGATGAGTGTACCCAAACCAAGTTAATGGGTCAAGGA 300
QY 301 ATTACCTGATGATCGACGCGTGCAAAAGGAGGAAATGTTAGCAAGTTCAGCAATG 360
DB 301 ATTACCTGATGATCGACGCGTGCAAAAGGAGGAAATGTTAGCAAGTTCAGCAATG 360
QY 361 TTGGGATATACATATGGAAGCTCATGGAAGAAAGCAATATTCGGCTCCAAATATGACAA 420
DB 361 TTGGGATATACATATGGAAGCTCATGGAAGAAAGCAATATTCGGCTCCAAATATGACAA 420
```

```
QY 421 CCAATGACGAAGATTATTGTGCCAGAGATCTTACGTTATGAGCAGACCATGTAC 480
    |||||||
Db 421 CCAATGACGAAGATTATTGTGCCAGAGATCTTACGTTATGAGCAGACCATGTAC 480
QY 481 GGCAGTGGCTGGAGTGGGCGTGAAGAGATGTCTTCCAGACGTTGATCTTGT 540
    |||||||
Db 481 GGCAGTGGCTGGAGTGGGCGTGAAGAGATGTCTTCCAGACGTTGATCTTGT 540
QY 541 TCCAGAACTTGATGGGAAAGAGTGTGTAAATGACCAAAAGATGACTTCCAGAGCTCA 600
    |||||||
Db 541 TCCAGAACTTGATGGGAAAGAGTGTGTAAATGACCAAAAGATGACTTCCAGAGCTCA 600
QY 601 CGCCGAGCTTAAACGAGATATCTCTCTGACACCTTACCTTACCTTACAGAGAGAGAG 660
    |||||||
Db 601 CGCCGAGCTTAAACGAGATATCTCTCTGACACCTTACCTTACCTTACAGAGAGAGAG 660
QY 661 CCACTTTTATTTTCCAAAATACATCAGTTTACCAGAGCAAGCAAGAAATTAACAACAA 720
    |||||||
Db 661 CCACTTTTATTTTCCAAAATACATCAGTTTACCAGAGCAAGCAAGAAATTAACAACAA 720
QY 721 GGCAGATTTTACCTTATGAGCAAGCAGAGAGATCAGCGTGGAGAGTCAACAGCATCCCA 780
    |||||||
Db 721 GGCAGATTTTACCTTATGAGCAAGCAGAGAGATCAGCGTGGAGAGTCAACAGCATCCCA 780
QY 781 CTCAGTCAAAAAGCTACCCAAACATCATCTTCAACAGTGGCCCAAAAGAGAGAGAGCTC 840
    |||||||
Db 781 CTCAGTCAAAAAGCTACCCAAACATCATCTTCAACAGTGGCCCAAAAGAGAGAGAGCTC 840
QY 841 CTCAGTTAGATCTCTTATGAGATCTTGGACCGACGACGACGCTCTTGGAAATCCAGAGGA 900
    |||||||
Db 841 CTCAGTTAGATCTCTTATGAGATCTTGGACCGACGACGACGCTCTTGGAAATCCAGAGGA 900
QY 901 GTGGGAGATATACAGCTATGAGCACTTCTGAGCTTCTGTGCGAGAGCTTCCAACTCCA 960
    |||||||
Db 901 GTGGGAGATATACAGCTATGAGCACTTCTGAGCTTCTGTGCGAGAGCTTCCAACTCCA 960
QY 961 ACTGCACTACCTGGGAGGAGCAAAATGGGAGTTCAAGATGACAGACCTGATGAAGTGG 1020
    |||||||
Db 961 ACTGCACTACCTGGGAGGAGCAAAATGGGAGTTCAAGATGACAGACCTGATGAAGTGG 1020
QY 1021 CTCGGGCTTGGGAGAGAGAGAAAGCAAACTTAACATGAACTATGACAACTCAGCCGTG 1080
    |||||||
Db 1021 CTCGGGCTTGGGAGAGAGAGAAAGCAAACTTAACATGAACTATGACAACTCAGCCGTG 1080
QY 1081 CACTTGCTACTACTATGACAAAAATATTATGACTAAAGTTTCAATGGTAAACGCTATGCT 1140
    |||||||
Db 1081 CACTTGCTACTACTATGACAAAAATATTATGACTAAAGTTTCAATGGTAAACGCTATGCT 1140
QY 1141 ACAAAATTTGATTTCCAGGAAATGGCTCAGAGCCCTCCAGCTCCAGAAATCATCCA 1200
    |||||||
Db 1141 ACAAAATTTGATTTCCAGGAAATGGCTCAGAGCCCTCCAGCTCCAGAAATCATCCA 1200
QY 1201 TGTACAAATATCCATCAGAGATCTCCCTACATGAGTTCTTACATGACACCCCGAGAGA 1260
    |||||||
Db 1201 TGTACAAATATCCATCAGAGATCTCCCTACATGAGTTCTTACATGACACCCCGAGAGA 1260
QY 1261 TGAACCTTTGAGCTGCCCATCCCTGCTTGGCCGTAACCTTATCAGAGTTTGTGCTG 1320
    |||||||
Db 1261 TGAACCTTTGAGCTGCCCATCCCTGCTTGGCCGTAACCTTATCAGAGTTTGTGCTG 1320
QY 1321 CCCCTAATCATCTGGAATTCACCAACTGGAGGATCTTACCCCAATACAGAGCTCCAG 1380
    |||||||
Db 1321 CCCCTAATCATCTGGAATTCACCAACTGGAGGATCTTACCCCAATACAGAGCTCCAG 1380
QY 1381 CTGCTATATGCTTCCATCTTTGGCACTTACTTAAGTGGGAAAGAAAGAGAGGCT 1440
    |||||||
Db 1381 CTGCTATATGCTTCCATCTTTGGCACTTACTTAAGTGGGAAAGAAAGAGAGGCT 1440
QY 1441 AAGAAAA 1447
    |||||||
Db 1441 AAGAAAA 1447
```

```
RESULT 2
US-08-878-177-3
; Sequence 3, Application US/08878177
; Patent No. 6294354
; GENERAL INFORMATION:
; APPLICANT: Chugai Selyaku Kabushiki Kaisha, 5001, Iwamoto et
; TITLE OF INVENTION: Cell Calcification Suppressing Proteins and Genes of
; FILE REFERENCE: Chugai Selyaku Kabushiki Kaisha 5001
; CURRENT APPLICATION NUMBER: US/08/878,177
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1528
; TYPE: DNA
; ORGANISM: c-ery gene, chicken DNA
US-08-878-177-3

Query Match          93.7%; Score 1356; DB 4; Length 1528;
Best Local Similarity 94.7%; Pred. No. 0;
Matches 1447; Conservative 0; Mismatches 0; Indels 81; Gaps 1;

QY 1 GAATTCGCGAAGCAATATTTATTTAGCAATTTATGCGATCAATATCTTGATCACA 60
    |||||||
Db 1 GAATTCGCGAAGCAATATTTATTTAGCAATTTATGCGATCAATATCTTGATCACA 60
QY 61 TTATGGCAAGCACTATTTAAGAAAGCATTTATCAGTGGAGTGAAGACAGTCTGTTTG 120
    |||||||
Db 61 TTATGGCAAGCACTATTTAAGAAAGCATTTATCAGTGGAGTGAAGACAGTCTGTTTG 120
QY 121 AGTGTGCTTACGAGATGCCCCACCTTGCAAAAGACAGAAATGACAGCTCTCTTCCAGTG 180
    |||||||
Db 121 AGTGTGCTTACGAGATGCCCCACCTTGCAAAAGACAGAAATGACAGCTCTCTTCCAGTG 180
QY 181 AATATGGCAAAATCATCAAGATGAGCCCGGCTTCCCGACGAGACTGGTTATCAGAC 240
    |||||||
Db 181 AATATGGCAAAATCATCAAGATGAGCCCGGCTTCCCGACGAGACTGGTTATCAGAC 240
QY 241 CCCCAGCAGAGTTTACCATTAAGATGAGTGAAGTAAACCAACAGGTTAATGGCTCAAGGA 300
    |||||||
Db 241 CCCCAGCAGAGTTTACCATTAAGATGAGTGAAGTAAACCAACAGGTTAATGGCTCAAGGA 300
QY 301 ATTACCTGATGACTCAGCGTGGCAAAAGAGGAAATGTTTACGATTTAGACAAATG 360
    |||||||
Db 301 ATTACCTGATGACTCAGCGTGGCAAAAGAGGAAATGTTTACGATTTAGACAAATG 360
QY 361 TTGGGATGAACCTATGAGAGTACATGGAAGAGAAAGCATATTCGGCTCCAAATATGACAA 420
    |||||||
Db 361 TTGGGATGAACCTATGAGAGTACATGGAAGAGAAAGCATATTCGGCTCCAAATATGACAA 420
QY 421 CCAATGAACGAAGATTATTTGTCAGCAGAGATCTTACGTTATGGAGCAGACCATGTAC 480
    |||||||
Db 421 CCAATGAACGAAGATTATTTGTCAGCAGAGATCTTACGTTATGGAGCAGACCATGTAC 480
QY 481 GGCAGTGGCTGGAGTGGGCGAGTGAAGAGATGTCTTCCAGACGTTGATCTTGT 540
    |||||||
Db 481 GGCAGTGGCTGGAGTGGGCGAGTGAAGAGATGTCTTCCAGACGTTGATCTTGT 540
QY 541 TCCAGAACTTGATGGGAAAGAGTGTGTAAATGACCAAAAGATGACTTCCAGAGCTCA 600
    |||||||
Db 541 TCCAGAACTTGATGGGAAAGAGTGTGTAAATGACCAAAAGATGACTTCCAGAGCTCA 600
QY 601 CGCCGAGCTTAAACGAGATATCTCTCTGACACCTTACCTTACCTTACAGAGAGAGAG 660
    |||||||
Db 601 CGCCGAGCTTAAACGAGATATCTCTCTGACACCTTACCTTACCTTACAGAGAGAGAG 660
QY 655 -----GAGAGGCCACTTTTATTTTTCCAAAATACATCAGTTTACCCAGAAAG 699
    |||||||
Db 655 -----GAGAGGCCACTTTTATTTTTCCAAAATACATCAGTTTACCCAGAAAG 699
```

```

Db 721 ATGCTAGAAAACAGAGGAGCCACTTTATTTTCCAAAATCAATGATTACCCAGAG 780
QY 700 CAACGCAAGAAATTAACAACAAGCCAGATTTACCTTATGAGCAAGAGAGATCAGCGT 759
Db 781 CAACGCAAGAAATTAACAACAAGCCAGATTTACCTTATGAGCAAGAGAGATCAGCGT 840
QY 760 GGAGAGTACAGCCATCCCACTAGTCAAAAGCTACCCACATCATCTTCAACAGTGC 819
Db 841 GGAGAGTACAGCCATCCCACTAGTCAAAAGCTACCCACATCATCTTCAACAGTGC 900
QY 820 CCAAAACAGAAAGACAGCGTCTCTAGTTAGATCTTATCAATCTTGGACCGACAGCA 879
Db 901 CCAAAACAGAAAGACAGCGTCTCTAGTTAGATCTTATCAATCTTGGACCGACAGCA 960
QY 880 GCCGCTTGGCAATCCAGAGGAGTGGGACAGATACAGTATGAGCAGTTCCTACTGAGCTTC 939
Db 961 GCCGCTTGGCAATCCAGAGGAGTGGGACAGATACAGTATGAGCAGTTCCTACTGAGCTTC 1020
QY 940 TGTGGACAGCTCCAACTCCAACTGCATCACCTGGAGGGACCAAAATGGGAGTTCAAGA 999
Db 1021 TGTGGACAGCTCCAACTCCAACTGCATCACCTGGAGGGACCAAAATGGGAGTTCAAGA 1080
QY 1000 TGAAGAACCTCGATGAAGTGGCTGGGGTGGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1059
Db 1081 TGACAGACCTCGATGAAGTGGCTGGGGTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
QY 1060 ACTATGACAAACTCAGCGCTGCTACTGCTACTATGACAAATATTTAGACTAAAG 1119
Db 1141 ACTATGACAAACTCAGCGCTGCTACTGCTACTATGACAAATATTTAGACTAAAG 1200
QY 1120 TTCATGTTAAAGCTATGCTTACCAATTTTGAATTTTCAAGGAATGCTCAGGCGCTCAGC 1179
Db 1201 TTCATGTTAAAGCTATGCTTACCAATTTTGAATTTTCAAGGAATGCTCAGGCGCTCAGC 1260
QY 1180 CTCACCTTCCAGATTCATTCATGTAACAATACCATCAGACCTCCCTACATGAGTTCCT 1239
Db 1261 CTCACCTTCCAGATTCATTCATGTAACAATACCATCAGACCTCCCTACATGAGTTCCT 1320
QY 1240 ACCATGACACACCCAGAAAGATGAATTTTGTAGCTCCCATCCCGTCTTGGCCGTA 1299
Db 1321 ACCATGACACACCCAGAAAGATGAATTTTGTAGCTCCCATCCCGTCTTGGCCGTA 1380
QY 1300 CCTCATCAGCTTTTGTGCTGCCCCATATCCATCTGGAATTCACCAATGAGAGCAATCT 1359
Db 1381 CCTCATCAGCTTTTGTGCTGCCCCATATCCATCTGGAATTCACCAATGAGAGCAATCT 1440
QY 1360 ACCCAATACAGGCTGCGAGCTGCTCATATGCTTCCCATCTTGGCACTACTACTAG 1419
Db 1441 ACCCAATACAGGCTGCGAGCTGCTCATATGCTTCCCATCTTGGCACTACTACTAG 1500
QY 1420 TGGGGAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1447
Db 1501 TGGGGAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1528

RESULT 3
US-08-343-443B-3
: Sequence 3, Application US/0834443B
: Patent No. 5968734
: GENERAL INFORMATION:
: APPLICANT: Aurias, Alain
: APPLICANT: Delattre, Olivier
: APPLICANT: Desmazes, Chantal
: APPLICANT: Melot, Thomas
: APPLICANT: Peter, Martine
: APPLICANT: Plooungastel, Beatrice
: APPLICANT: Thomas, Gilles
: APPLICANT: Zugman, Jessica
: TITLE OF INVENTION: NUCLEIC ACID CORRESPONDING TO A GENE OF
: TITLE OF INVENTION: CHROMOSOME 22 INVOLVED IN RECURRENT CHROMOSOMAL
: TITLE OF INVENTION: TRANSLATIONS ASSOCIATED WITH THE DEVELOPMENT OF CANCEROUS
: TITLE OF INVENTION: TUMORS, AND NUCLEIC ACIDS OF FUSION RESULTING FROM SAID

```

```

: TITLE OF INVENTION: TRANSLOCATIONS
: NUMBER OF SEQUENCES: 129
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Weiser & Associates
: STREET: 230 South Fifteenth Street
: CITY: Philadelphia
: STATE: PA
: COUNTRY: USA
: ZIP: 19102
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: AEDIT 1.0 DOS text editor
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/343,443B
: FILING DATE: 18-NOV-1994
: CLASSIFICATION: 514
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/FR93/00494
: FILING DATE: 19-MAY-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: FR 92/06123
: FILING DATE: 20-MAY-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Weiser, Gerard J.
: REGISTRATION NUMBER: 19,763
: REFERENCE/DOCKET NUMBER: 989.6121P
: TELEPHONE: 215-875-8383
: TELEFAX: 215-875-8394
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2938 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 143..1498
: US-08-343-443B-3

Query Match 39.0%; Score 563.8; DB 2; Length 2938;
Best Local Similarity 66.5%; Pred. No. 4.4e-184;
Matches 913; Conservative 0; Mismatches 427; Indels 33; Gaps 6;

```

```
QY 477 GTACGGCAATGCTGAGATGGGAGTGAAGAGTATGCTTCTCCAGACGTGACATCTTG 536
    || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 554 GTGAGGCAATGGCTGAGTGGGGCCATAAAGAGTATAGCTTGATGAGATCGACACATCC 613
QY 537 TTGTTCCAAACATATGATGGAAAGGTGTGTAATAATACCAAAATGATTAATCCAGAG 596
    || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 614 TTTTTCACAAACATGATGATGAGAGTGTGTAATAATGAAAGAGGAGGCTTCTCCCG 673
QY 597 CTCACGCCAGCATATATACGCAGATATCTCTCTACACCTACCTACCTACCTACAGAGAGA 656
    || || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 674 GCCACACCCCTCTACAAACAGAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 732
QY 657 GGAAGCCATTTTATTTTCCAAATACATCATGTTTACCCAGAACCAAGCAAGATATACA 716
    || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 733 -----TTCACTGCTGCTATATATCAACCTCCACACCCAGCAATCTCACATTTGAGT 787
QY 717 ACAAGGCCAGATTTTACCTATGATGAGAGAGAGAGATGCGGTGAGAGTCCAGCCAT 776
    || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 788 GTCAAAAGAAAGACCCCTTCTTATGATGATGATGATGATGATGATGATGATGATGATG 847
QY 777 CC---CACTCAGTCAAAAGCTACCAACCATCATCTTCAACAGTGGCCAAACAGAGAG 833
    || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 848 TCTGGCCCTCAACAAAGCTCTCCCTTGGAGGGGACAAACAGATCATAGATATACAGAG 907
QY 834 CAGCGTCTCATGTTAGATCTTATGATGATGATGATGATGATGATGATGATGATGATG 893
    || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 908 CAACGCCCCCAACAGATCTGATGATGATGATGATGATGATGATGATGATGATGATG 967
QY 894 CCAGGAGTGGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 953
    || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 968 CTTGGAAGGGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1027
QY 954 AACTCAACTGATCATCTGAGGAGGACCAAAATGGGAGTTCAAGTACAGACACCTGAT 1013
    || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 1028 AACGCCAGTGTATCATCTGAGGAGGGGACCAAGGGGAGTTCAAAATGAGAGAGCCGAT 1087
QY 1014 GAAAGTGTCTGGGCTTGGGAGAGAGAGAAACCTTAACATGATGATGATGATGATG 1073
    || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 1088 GAGGTGCGCAGGCGCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1147
QY 1074 ACCCGTCACTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1133
    || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 1148 ACCCGGCGCTCGTATATGATGATGATGATGATGATGATGATGATGATGATGATG 1207
QY 1134 TATGCTTCAAAATTTGATTTGACAGGAATGCTCAGGCGCTCCAGCCTCAGCAG 1193
    || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 1208 TATGCTTCAAAATTTGATTTGACAGGATTTGCGAGGCTCTGAGGACATCGAG 1267
QY 1194 TCATCCATGTAATAATACCATCAAGACCTCCCTTACATGATGATGATGATGATG 1253
    || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 1268 TGTGCTCATGTAAGTACCTCTTGTGATCTCTTACATGCTCTTACATGCTCCACAG 1327
QY 1254 CAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1313
    || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 1328 CAGAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1387
QY 1314 TTTTGTGCTGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1371
    || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 1388 TTTTGTGCTGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1447
QY 1372 -----GGCTGCGAGCTGCTCATATGATGATGATGATGATGATGATGATGATG 1417
    || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 1448 GTTCCCGGCTATCTTAACACCGACGCTTCACTTACAGGAGTATGAGGAGTATG 1500
```

RESULT 4
US-09-360-779-1

; Sequence 1, Application US/09360779
; Patent No. 6268216
; GENERAL INFORMATION:
; APPLICANT: Denieris, Evan S.
; APPLICANT: Eyodoro, Dmitry V.

```
; APPLICANT: Hendricks, Timothy J.  
; TITLE OF INVENTION: Reagents and Methods for the Screening of Compounds  
; FILE REFERENCE: CASE-03828  
; CURRENT APPLICATION NUMBER: US/09/360,779  
; EARLIER FILING DATE: 1999-07-26  
; EARLIER FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 13  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 1752  
; TYPE: DNA  
; ORGANISM: Rattus norvegicus  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (112)..(1131)  
US-09-360-779-1  
Query Match 12.4%; Score 180; DB 4; Length 1752;  
Best Local Similarity 74.0%; Pred. No. 1.5e-51;  
Matches 228; Conservative 0; Mismatches 80; Indels 0; Gaps 0;
```

```
QY 879 AGCGCTTGCAAAATCCAGAGAGTGGGAGATACAGTATGAGTCTTCTACTGAGCTT 938
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 529 AGCGCTGCGGTACAGAAAGGAGGAGAGATTCAGTTGTGAGTCTTCTACTGAGCTG 588
QY 939 CTGTCGAGAGCTGCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACT 998
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 589 CTGCGAGACCGCGGAGAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 648
QY 999 ATGACAGACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1058
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 649 CTCACGAGACCGGAGAGAGTGGGCGACCTGGGCGGAGCGCAAGCGCAAGCCATATG 708
QY 1059 AACTATGACAACTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1118
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 709 AACTATGACAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 768
QY 1119 GTTCAATGTAAGCGTATGATGATGATGATGATGATGATGATGATGATGATGATG 1178
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 769 GTGACAGGAGAGGCGTACGCTTGTGACTTCCAGAGGCTGCGACAGGCTTGCCAG 828
QY 1179 CCTGACCC 1186
    || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 829 CCACGACG 836
```

RESULT 5

US-09-435-335-1
; Sequence 1, Application US/09435335
; Patent No. 6384204

; GENERAL INFORMATION:
; APPLICANT: Denieris, Evan S.
; APPLICANT: Eyodoro, Dmitry V.
; APPLICANT: Hendricks, Timothy J.
; TITLE OF INVENTION: Reagents and Methods for the Screening of Compounds
; FILE REFERENCE: CASE-04027
; CURRENT APPLICATION NUMBER: US/09/435,335
; CURRENT FILING DATE: 1999-11-05
; EARLIER FILING DATE: 1999-07-26
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1752
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (112)..(1131)
US-09-435-335-1

QY 790 AACCTACCCACCATCATCTTCAACAGTGGCCCAAGACAGACGCTTCTCATTAG 849
DB 125 GTCTCCCTTGGAGGGGACCAACGATCATAGATATACAGACAGCCCGCCAGCCAG 184
QY 850 ATCTTATCATGATTTCTTGACCGACGACCGCTTTCGAATTCGAGGAGTGGGAGA 909
DB 185 ATCCGATCATGATCTTGGGCGCCAGCAGCGTGGCTTACGCAACCTGGAGAGGGGAGA 244
QY 910 TACAGTATGAGCAGTCTTCTACTGAGCTTGTGCGACAGCTCCACTCCACTGATCA 969
DB 245 TCCAGCTGGGCAATCTCTCTGAGAGTGTCTCCGACAGCCCAACGCGAGCTGATCA 304
QY 970 CCTGGAGGGCACAATGGGAGT 993
DB 305 CCTGGAGGGGACCAAGGGGAGT 328

RESULT 12

US-08-875-944B-1
; Sequence 1, Application US/08875944B
; Patent No. 6096542
; GENERAL INFORMATION:
; APPLICANT: FUJINAGA, Kei
; APPLICANT: YOSHIDA, Koichi
; APPLICANT: HICASHINO, Fumihito
; TITLE OF INVENTION: CANCER CONTROL
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.
; STREET: 624 Ninth Street N.W., Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/875,944B
; FILING DATE: 07-AUG-1997
; CLASSIFICATION: 514
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: JP 07-020173
; FILING DATE: 08-FEB-1995
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP96/00016
; FILING DATE: 09-JAN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWDY, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: FUJINAGA-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-5197
; TELEFAX: (202) 737-3528
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2064 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1386
; US-08-875-944B-1

Query Match 6.6%; Score 96; DB 3; Length 2064;
Best Local Similarity 62.5%; Pred. No. 1.8e-22;
Matches 150; Conservative 0; Mismatches 90; Indels 0; Gaps 0;
QY 910 TACAGTATGAGCAGTCTTCTACTGAGAGTGTCTGCGACAGCTCCAACTCCACTGATCA 969

DB 956 TGCAGCTGTGGCAATTTCTGTGGCTTGTCTGATGACCCCAAAAAGCCCATTTTCATGG 1015
QY 970 CCTGGAGGGCACAATAATGGGAGTTCAAGATGACAGACCTGTGAAGTGGCTGGGCTT 1029
DB 1016 CTTGAGAGGGCCGGGGAATGAGATGATCAAGCTCATTTAGCTTGAGAGGTGCCAGGCTCT 1075
QY 1030 GGGGAGAGGAGAAAAGCAACCTTAACATGACATGACAAACCTAGCCGTGCACTTGGCT 1089
DB 1076 GGGGCAATCCAGAAAGACCGGCGCAGCATGAAATTACGACAGCTGAGCGCGCTCCGAT 1135
QY 1090 ACTACTATGACAAATAATATTATGACTAAAGTTGATGTAAGCTTATGCTTACAAATTTG 1149
DB 1136 ACTATTATGAGAAAGCATCATGACAGAGGTGGCTGTGAGCGTTACGTGTAAGTTTG 1195

RESULT 13

US-09-116-049-3
; Sequence 3, Application US/09116049A
; Patent No. 6248351
; GENERAL INFORMATION:
; APPLICANT: Hung, Men-Chie
; TITLE OF INVENTION: HUMAN PEA3 IS A TUMOR SUPPRESSOR FOR CANCER CELLS
; FILE REFERENCE: UTSC:582
; CURRENT APPLICATION NUMBER: US/09/116,049A
; CURRENT FILING DATE: 1998-07-15
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 3
; LENGTH: 2064
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-116-049-3

Query Match 6.6%; Score 96; DB 4; Length 2064;
Best Local Similarity 62.5%; Pred. No. 1.8e-22;

Matches 150; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

QY 910 TACAGTATGAGCAGTCTTCTACTGAGCTTGTGCGACAGCTCCAACTCCACTGATCA 969
DB 956 TGCAGCTGTGGCAATTTCTGTGGCTTGTCTGATGACCCCAAAAAGCCCATTTTCATGG 1015
QY 970 CCTGGAGGGCACAATAATGGGAGTTCAAGATGACAGACCTGTGAAGTGGCTGGGCTT 1029
DB 1016 CTTGAGAGGGCCGGGGAATGAGATGATCAAGCTCATTTAGCTTGAGAGGTGCCAGGCTCT 1075
QY 1030 GGGGAGAGGAGAAAAGCAACCTTAACATGACATGACAAACCTAGCCGTGCACTTGGCT 1089
DB 1076 GGGGCAATCCAGAAAGACCGGCGCAGCATGAAATTACGACAGCTGAGCGCGCTCCGAT 1135
QY 1090 ACTACTATGACAAATAATATTATGACTAAAGTTGATGTAAGCTTATGCTTACAAATTTG 1149
DB 1136 ACTATTATGAGAAAGCATCATGACAGAGGTGGCTGTGAGCGTTACGTGTAAGTTTG 1195

RESULT 14

US-08-780-835B-1
; Sequence 1, Application US/08780835B
; Patent No. 5922688
; GENERAL INFORMATION:
; APPLICANT: Hung, Men-Chie
; APPLICANT: Xing, Xiangming
; TITLE OF INVENTION: PEA3 Is a Tumor Suppressor
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ARNOLD, WHITE AND DUNKER
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210-4433

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/780,835B
FILING DATE: 10-JAN-1997
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Wilson, Mark B.
REGISTRATION NUMBER: 37,259
REFERENCE/DOCKET NUMBER: UTSC500
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (512) 474-7577
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2410 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-780-835B-1

Query Match 6.6%; Score 95.2; DB 2; Length 2410;

Best Local Similarity 57.9%; Pred. No. 3.9e-22;
Matches 169; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

QY 865 TTGACCGACGACGACGCGCTTCAATCCAGGAGTGGCAGATACAGCTATGGCAGT 924
DB 1315 TTGAGCTTTCCGGAGGGGCGCACCTACGACGCGCGGGGCTTCAACTGTGGCAGT 1374
QY 925 TCTTACTGAGCTTCTGTGCGACAGCTCCAACTCCAACTGCATACCTGGAGGGCACAA 984
DB 1375 TTTGGTGGGCGCTCTGTGATACCCCAACAAATGCTCATTTGCTTGGACAGCGCGG 1434
QY 985 ATGGGAGTTCAAGTACAGACCCGTGTAAGTGGCTCGGCTTGGGGAGAGAGAAA 1044
DB 1435 GAATGAGTTTAACTAATTAATGACCTGAGAGGTTGCCAGCTCTGGGGATCCAGAA 1494
QY 1045 GCAACCTAATACATGATGACAACTCAACCTGACCTGCTGCTACTACTATGACAAA 1104
DB 1495 ACCGGCCAGCCATGATTAATGACAAAGTGAAGCGCGCTGCTGCTGATTAATGAGAA 1554
QY 1105 ATATTATGACTAAAGTTGATGTAAGCTTAAGCTTAAGCTTAAGCTTAAGCTTA 1156
DB 1555 GCATCATGACGAAGAGTGGCTGCGAAGCTGATGATGATGATGATGATGATGATG 1606

RESULT 15
US-09-303-268-1
Sequence 1, Application US/09303268
Patent No. 6172212

GENERAL INFORMATION:
APPLICANT: Hung, Mien-Chie
TITLE OF INVENTION: xing, xiangming
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: ARNOLD, WHITE AND DORKEE
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: USA
ZIP: 77210-4433
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/303,268
FILING DATE: 30-Apr-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/780,835
FILING DATE: 10-JAN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Wilson, Mark B.
REGISTRATION NUMBER: 37,259
REFERENCE/DOCKET NUMBER: UTSC500
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (512) 474-7577
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2410 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-303-268-1

Query Match 6.6%; Score 95.2; DB 4; Length 2410;

Best Local Similarity 57.9%; Pred. No. 3.9e-22;
Matches 169; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

QY 865 TTGACCGACGACGACGCGCTTCAATCCAGGAGTGGCAGATACAGCTATGGCAGT 924
DB 1315 TTGAGCTTTCCGGAGGGGCGCACCTACGACGCGCGGGGCTTCAACTGTGGCAGT 1374
QY 925 TCTTACTGAGCTTCTGTGCGACAGCTCCAACTCCAACTGCATACCTGGAGGGCACAA 984
DB 1375 TTTGGTGGGCGCTCTGTGATACCCCAACAAATGCTCATTTGCTTGGACAGCGCGG 1434
QY 985 ATGGGAGTTCAAGTACAGACCCGTGTAAGTGGCTCGGCTTGGGGAGAGAGAAA 1044
DB 1435 GAATGAGTTTAACTAATTAATGACCTGAGAGGTTGCCAGCTCTGGGGATCCAGAA 1494
QY 1045 GCAACCTAATACATGATGACAACTCAACCTGACCTGCTGCTACTACTATGACAAA 1104
DB 1495 ACCGGCCAGCCATGATTAATGACAAAGTGAAGCGCGCTGCTGCTGATTAATGAGAA 1554
QY 1105 ATATTATGACTAAAGTTGATGTAAGCTTAAGCTTAAGCTTAAGCTTAAGCTTA 1156
DB 1555 GCATCATGACGAAGAGTGGCTGCGAAGCTGATGATGATGATGATGATGATGATG 1606

Search completed: November 9, 2002, 12:29:48
Job time: 113.808 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 9, 2002, 16:45:07 ; Search time 55.8288 Seconds
(without alignments)
1664.502 Million cell updates/sec

Title: US-09-902-772-2

Perfect score: 2444
Sequence: 1 MASTIKREALSVSEDSLF.....IYPNRLPAHMPSHLGITY 451

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: SP:archaea:*
2: SP:bacteria:*
3: SP:fungi:*
4: SP:human:*
5: SP:invertebrate:*
6: SP:mammal:*
7: SP:mnc:*
8: SP:organelle:*
9: SP:phage:*
10: SP:plant:*
11: SP:rodent:*
12: SP:virus:*
13: SP:vertebrate:*
14: SP:unclassified:*
15: SP:rvirus:*
16: SP:bacteriap:*
17: SP:archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2444	100.0	451	13	08UUN0	08UUN0 gallus gall
2	2267	92.8	486	11	0920K7	0920K7 mus musculu
3	2192	89.7	456	13	09W6Z9	09W6Z9 xenopus lae
4	2179	89.2	455	11	091XV5	091XV5 ratius norv
5	2159.5	88.4	463	11	0920K8	0920K8 mus musculu
6	2153.5	88.1	485	13	09W700	09W700 xenopus lae
7	2151	88.0	451	11	0920K9	0920K9 mus musculu
8	1637	67.0	451	13	09PU61	09PU61 brachydanto
9	1579.5	64.6	432	13	0934Z5	0934Z5 coturnix co
10	1486.5	60.8	414	13	09PUL6	09PUL6 brachydanto
11	965	39.5	476	4	09BBD1	09BBD1 homo sapien
12	661	27.0	257	5	095162	095162 drosophila
13	586.5	24.0	164	5	095ZEO	095ZEO hediste div
14	543.5	22.2	160	5	09N9D6	09N9D6 perinereis
15	524	21.4	377	5	022355	022355 caenorhabd1
16	508.5	20.8	238	4	099581	099581 homo sapien

17	507	20.7	237	11	08Q2W2	08Q2W2 mus musculu
18	507	20.7	340	11	070132	070132 ratius norv
19	477	19.5	208	5	018579	018579 caenorhabd1
20	460.5	18.8	336	6	062804	062804 ovis aries
21	457	18.7	351	11	091YX8	091YX8 mus musculu
22	448	18.3	393	11	0921D8	0921D8 mus musculu
23	440	18.0	455	13	0919E8	0919E8 brachydanto
24	434	17.8	559	5	026645	026645 strongyloce
25	432.5	17.7	426	6	062803	062803 ovis aries
26	428.5	17.5	440	13	09YHUS	09YHUS fuqu rubrip
27	420.5	17.2	468	11	099K36	099K36 mus musculu
28	341	14.0	268	13	091744	091744 xenopus lae
29	329.5	13.5	477	13	09YHW6	09YHW6 gallus gall
30	325.5	13.3	613	5	09VC38	09VC38 drosophila
31	324	13.3	513	11	090ZM1	090ZM1 mus musculu
32	324	13.3	513	11	08R4Z4	08R4Z4 mus musculu
33	322	13.2	512	4	09BX30	09BX30 homo sapien
34	321	13.1	161	5	095ZEL	095ZEL hediste div
35	320	13.1	115	5	0967C5	0967C5 hediste div
36	320	13.1	115	5	09W696	09W696 xenopus lae
37	316	12.9	477	13	093322	093322 xenopus lae
38	315.5	12.9	441	5	027378	027378 caenorhabd1
39	311	12.7	401	5	0965U5	0965U5 caenorhabd1
40	310	12.7	491	13	09PUQ1	09PUQ1 brachydanto
41	307.5	12.6	494	13	057586	057586 brachydanto
42	306.5	12.5	465	13	090ZS8	090ZS8 fuqu rubrip
43	303.5	12.4	444	13	090ZS9	090ZS9 brachydanto
44	303	12.4	325	11	09WTP3	09WTP3 mus musculu
45	300.5	12.3	143	4	087AC8	087AC8 homo sapien

ALIGNMENTS

RESULT 1	ID	DB	PRELIMINARY	PRT	451 AA.
08UUN0	08UUN0	08UUN0	08UUN0		
AC	08UUN0	08UUN0	08UUN0		
DT	01-MAR-2002 (TREMBLrel. 20, Created)				
DT	01-MAR-2002 (TREMBLrel. 20, Last sequence update)				
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)				
DE	Erg Isoform C-1-1.				
OS	Gallus gallus (Chicken).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;				
OC	Gallus.				
OX	NCBI_TaxID=9031;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	MEDLINE-20351415; Pubmed-10893254;				
RX	Iwamoto M., Higuchi Y., Enomoto-Iwamoto M., Kurisu K.,				
RA	Yeh H., Rosenblom J., Pacifici M.,				
RT	"The role of ERG (ets related gene) in cartilage development.";				
RL	Osteoarthritis Cartilage 9:S41-S47(2001).				
DR	EMBL: AY065661; AAL40889.1; -				
DR	InterPro: IPR000418; Ets.				
DR	InterPro: IPR002341; HSF-ETS.				
DR	InterPro: IPR001660; SAM.				
DR	InterPro: IPR003118; SAM_PNT.				
DR	Pfam: PF001178; Ets; 1.				
DR	Pfam: PF02198; SAM_PNT; 1.				
DR	PRINTS: PR00454; ETSDOMAIN.				
DR	SMART: SM00413; ETS; 1.				
DR	SMART: SM00251; SAM_PNT; 1.				
DR	PROSITE: PS00345; ETS_DOMAIN_1; UNKNOWN_1.				

	Best Local Similarity	89.9%; Pred. No. 2,3e-165;			
	Matches	402; Conservative	21; Mismatches	22; Indels	2; Gaps
OY	5	IKEALSVSEDSOLFECAYGSPHLAKTENTASSSSSEYGQTSKMSPRVPQDMLSOPARY	64		
Db	12	IKEALSVSEDSOLFECTGYGTPLTKTEMTASSSDPYGQTSKMSPRPVQDMLSOPRSRY	71		
OY	65	TIKMCNPNQNVNGSRNSPDSCVAAGGKKGVSSSDVNGMYGYSMEKHLPNNMTTNER	124		
Db	72	TIKMCNPNQNVNGSRNSPDSCISIGKSGKGGSDVNGMYGSYIEKHISPNNMTTNER	131		
OY	125	VIVPADPTLMSIDHROWLEMAVKFEGLDVDVILLFONTIDGELCKMTKDDEORTLPSVN	184		
Db	132	VIVPADPTLMSIDHROWLEMAIKKEYGLDVVLFLQNIDGELCKMTKEDFHRLPSTN	191		
OY	185	ADILLSHLYLRERGAATFEPMYSVEAQTITTRPDLPIYAQRSAMTSSHSPQSA	244		
Db	192	ADILLSHLHYLRERGAATFEPMYSVYQADNQIIPSNODLSYEPSRSAMTNHPAP - PSKA	250		
OY	245	TOPSSTVPKTEDORPDDPVQIILPPTSRLANPGSGOIQLOMFTLELLSDSSNSNCITW	304		
Db	251	SOP-STVPKTEDPPOLDPPYOILPPTSRLANPGSGOIQLMFTLELLSDSSNSNCITW	309		
OY	305	EETNEEFKTTDDEAVARMGGERKSKPNMYDKLSRALRYVYNKITMKHGKRYAKPDF	364		
Db	310	EETNEEFKTTDDEAVARKGERKSKPNMYDKLSRLRITYDNKITMTKGHRVAKPDF	369		
OY	365	HGIQAOLAPHPRESSMYKXPSDLPYMSSYHAHPDKMNFAPPPALPYVSSSFFAAPNPY	424		
Db	370	HGIQAOLAPHPRESIMYKXPSELPLPMSSYHAHPDKMFAPPPALPYVSSSFFAAPNPY	429		
OY	425	WNSPGTGIYPNTRLPAAMKPSHLGTYT	451		
Db	430	WNSPGTGIYPNTRLPASHSLSHLGTYT	456		
	RESULT 4				
	O91XVS				
ID	O91XVS	PRELIMINARY:	PRT:	455 AA.	
AC	O91XVS:				
DT	01-DEC-2001 (TREMBLrel. 19, Created)				
DT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)				
DT	01-MAR-2002 (TREMBLrel. 20, Last annotation update)				
DE	Vascular endothelial cell specific protein 14.				
GN	VESPL4.				
OS	Rattus norvegicus (Rat).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
CC	Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.				
OX	NCB1_TaxID=10116;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=LIVER.				
RA	Aoki T., Toyoda H., Nishimoto S., Tawara J., Urai Y., Komuraaki T.;				
RT	*Identification of VESPL4,a vascular endothelial cell specific				
	protein.?				
RL	Submitted (Aug-1999) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; AB031088; BAB62744.1; .				
DR	InterPro: IPR000418; Ets.				
DR	InterPro: IPR002341; HSF_Ets.				
DR	InterPro: IPR003118; SAM_PWT.				
DR	pfam; PF00178; Ets; 1.				
DR	pfam; PF00198; SAM_PWT; 1.				
DR	PROSITE; PS00345; ETS_DOMAIN_1; UNKNOWN_1.				
DR	PROSITE; PS00346; ETS_DOMAIN_2; UNKNOWN_1.				
DR	PROSITE; PS0061; ETS_DOMAIN_3; 1.				
SO	SEQUENCE 455 AA; 51382 MW; b/E2A1564F8B560D CRC64;				
	Query Match	89.2%; Score 2179; DB 11; Length 455;			
	Best Local Similarity	88.3%; Pred. No. 2.4e-164;			
	Matches	406; Conservative	14; Mismatches	26; Indels	14; Gaps
OY	1	MASTIKEALSVSEDSOLFECAYGSPHLAKTENTASSSSSEYGQTSKMSPRVPQDMLSOP	60		

ID	Q920K8	PRELIMINARY	PROT	463 AA
Db	1	MASTIKKEALSVSKQOSLFECAVGTPIHLAKTEMTASSSSSDYQOTSKMSPRVPOQDWLSQP	60	
Qy	61	PARVTIKMECNPNQVNGSRNSPDCCSVAKGKMWSSSDVNGVNGSYMEKHIIPENMTT	120	
Db	61	PARVTIKMECNPNQVNGSRNSPDCCSVAKGKMWSSSDVNGVNGSYMEKHIIPENMTT	120	
Qy	121	NERRIVPADPPLMNSTDHRQWLENAVKETGLPYVDVILLFQNIIDGKELCKMTKDFOLTL	180	
Db	121	NERRIVPADPPLMNSTDHRQWLENAVKETGLPYVDVILLFQNIIDGKELCKMTKDFOLTL	180	
Qy	181	PSYNDILLSHHYRERKATFIPIFNTSVYPAOTORITRR-----DLYEQARSA	232	
Db	181	PSYNDILLSHHYRERKATFIPIFNTSVYPAOTORITRR-----DLYEQARSA	232	
Qy	233	WTSHSHPT-QSKATPSSSTVPEKTEQDORPOLDPYOILGPTSSRLANPGSGQIQLOWFILE	291	
Db	236	WTGSHHPPTQSGAASPSPETVTKTEQDORQODPYQILGPTSSRLANPGSGQIQLOWFILE	295	
Qy	232	LLSDSSNSNCITWEGTNGEFTKMTDDEVARRWGERKSKPNMNYDKLSRLARYYDKNIMT	351	
Db	296	LLSDSSNSNCITWEGTNGEFTKMTDDEVARRWGERKSKPNMNYDKLSRLARYYDKNIMT	355	
Qy	332	KVHGKRYVAKPFHFGHIAOLOPHPRESSYKXKPSDLPYVSSVHAHPQKNFPAHPPALP	411	
Db	336	KVHGKRYVAKPFHFGHIAOLOPHPRESSYKXKPSDLPYVSSVHAHPQKNFPAHPPALP	415	
Qy	412	VTSSSFPAAPNPYNSPTGGIYPNTRLPAHMPSHLGTYY	451	
Db	416	VTSSSFPAAPNPYNSPTGGIYPNTRLPAHMPSHLGTYY	455	
RESULT 5				
ID	Q920K8	PRELIMINARY	PROT	463 AA
AC	Q920K8			
DT	01-DEC-2001 (TREMBLrel. 19, Created)			
DT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)			
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)			
DE	Erg protein.			
GN	Erg.			
OS	Mus musculus (Mouse).			
OC	Eukaryota, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Ozawa R., Noguchi H., Taylor T.D., Takeda T., Hattori M., Sakaki Y.;			
RT	"Mus musculus Erg mRNA."			
RL	Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.			
DR	EMBL; AB073079; BAB69949.1; -			
DR	MGD; MGI:95415; Erg.			
DR	InterPro; IPR000418; Ets.			
DR	InterPro; IPR003118; SAM_PNT.			
DR	Pfam; PF00178; Ets_1.			
DR	Pfam; PF02198; SAM_PNT; 1.			
DR	PROSITE; PS00345; ETS_DOMAIN_1; UNKNOWN_1.			
DR	PROSITE; PS00346; ETS_DOMAIN_2; UNKNOWN_1.			
DR	PROSITE; PS50061; ETS_DOMAIN_3; 1.			
DR	SEQUENCE 463 AA; 51971 MW; 510D2BB7663D4A4 CRC64;			
Qy	Query Match	88.4%; Score 2159.5; DB 11; Length 463;		
Db	Best Local Similarity	84.5%; Pred. No. 8.8e-163;		
Db	Matches 403; Conservative 12; Mismatches 13; Indels 49; Gaps 2;			
Qy	2	ASTIKKEALSVSEDSLFECAVGSPLAKTEMTASSSEYQGTSKMSPRVPOQDWLSQP	61	
Db	9	AAHIKEALSVSEDSLFECAVGSPLAKTEMTASSSEYQGTSKMSPRVPOQDWLSQP	68	
Qy	62	ARVTIKMECNPNQVNGSRNSPDCCSVAKGKMWSSSDVNGVNGSYMEKHIIPENMTT	121	
Db	69	ARVTIKMECNPNQVNGSRNSPDCCSVAKGKMWSSSDVNGVNGSYMEKHIIPENMTT	128	

QY	122	ERRATVPADPLMTSTGDHROWLEMAVEYEGJPDVLDLFONISKEKCKMTKDPQCLTP	181
Db	129	ERRATVPADPLMTSTGDHROWLEMAVEYEGJLDVDVLLFONISKEKCKMTKDPQCLTP	188
QY	182	SYNDLILSHLYRER-----GATFIPNPTSVYPEAT	214
Db	189	SYNDLILSHLYRERPLPLTSDVDKALONSPRLMHARNTGGAAFIIPNTSVYPEAT	248
QY	215	QRITTRPDLPEQARRSAMTSHSHPTOSKATOPSSSTVPKTEDORPOLDLYQILGPTSSR	274
Db	249	QRITTRP-----AAQPSAIVPKTEDORPOLDLYQILGPTSSR	286
QY	275	LANGSGOIOIMOFLELHLDSSNCTIWEGTGFEKMTDPDEVARMEERKSPPMNY	334
Db	287	LANGSGOIOIMOFLELHLDSSNCTIWEGTGFEKMTDPDEVARMEERKSPPMNY	346
QY	335	DKLSRALRYYYDKNIMTKVHKRRAAYKFDHGLAQALQAPHPRESSMYKYPDDLPLYMSSYH	394
Db	347	DKLSRALRYYYDKNIMTKVHKRRAAYKFDHGLAQALQAPHPRESSLYKYPDDLPLYMSSYH	406
QY	395	AHPQKMNVAHPALPVTSSSPFAAANPNWNSPTGGIYENTRLPAHMHPSHLGTY	451
Db	407	AHPQKMNVAHPALPVTSSSPFAAANPNWNSPTGGIYENTRLPAHMHPSHLGTY	463

RESULT 6	
Q9W700	
ID	Q9W700
PRELIMINARY;	PRT; 485 AA

DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Erg protein.
GN ERG.
OS *Xenopus laevis* (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae;
OC Xenopodidae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP
SEQUENCE FROM N.A.
RA MEDLINE=20099678; PubMed=10633861;
RX Balzinger M., Mager-Heckel A.M., Remy P.;
RT "Erg: Expression and pattern and overexpression during development
RT of a role in endothelial cell differentiation.",
RL Dev. Dyn. 216:420-433(1999).
DR EMBL: AJ224125; CAB4566.1; -.
DR HSSP: Q01543; IFL1.
DR InterPro: IPR000418; Ets.
DR InterPro: IPR002341; HSF_ETS.
DR InterPro: IPR003118; SAM_PNT.
DR Pfam: PR00178; Ets_1.
DR Pfam: PF02198; SAM_PNT; 1.
DR PRINTS: PR00454; ETSDOMAIN.
DR SMART: SM00413; ETS; 1.
DR SMART: SM00251; SAM_PNT; 1.
DR PROSITE: PS00345; ETS_DOMAIN_1; 1.
DR PROSITE: PS00346; ETS_DOMAIN_2; 1.
DR PROSITE: PS00061; ETS_DOMAIN_3; 1.
SQ
SEQUENCE 485 AA; 54804 MW; F87BF36DAFBA279F CRC64;

Query Match	88.1%	Score 2153.5	DB 13	Length 485
Best Local Similarity	84.5%	Pred. No. 2.8e-162		
Matches 399	Conservative 21	Mismatches 23	Indels 29	Gaps 3

0Y	EALSVSPDQSLFPCACVSGPLHATKEMTASSSEYGGTSMKSPVPQDMLSDPAPVTI	66
7	EALSVSPDQSLFPCACVSGPLHATKEMTASSSEYGGTSMKSPVPQDMLSDPAPVTI	75
16	EALSVSPDQSLFPCACVSGPLHATKEMTASSSEYGGTSMKSPVPQDMLSDPAPVTI	75
0Y	KMECNPNQVNGSRNSPDDCSYAKGCKGVSSSDNYGNNTGYSTMEKHITPPNMTTNERVTI	126
67	KMECNPNQVNGSRNSPDDCSYAKGCKGVSSSDNYGNNTGYSTMEKHITPPNMTTNERVTI	135
76	KMECNPNQVNGSRNSPDDCSYAKGCKGVSSSDNYGNNTGYSTMEKHITPPNMTTNERVTI	135

QY	127	VPAPRTLMSTDHVNQWEMAKKXEGELGDVYLLEFQNDIGKELCKMTDQRLPLPSNAD	186
Db	136	VPAPRTLMSTDHVNQWEMAKKXEGELGDVYLLEFQNDIGKELCKMTDQRLPLPSNAD	195
QY	187	ILLSHLNHLRER-----GATLPFPTSVYPEATORITTT	219
Db	196	ILLSHLNHLRPTPLPHLTSDDVDKALONSPRLMHARNTGASGFIFPNSVYOANOIRIP	255
QY	220	RPLDPEYEQARRSAMTSSHPQSKATQPSSTSVYKTEDEQRQOLDPYQLLPTSSRLANPG	279
Db	256	RQDLSYEPRSRSAMTNNHAP-PSKASQ-R-STYVYKTEDEPRQOLDPYQLLPTSSRLANPG	313
QY	280	SGQQLQMLQLLELLSDSSNSCITWBEETNGEFKMTDDEVARRRNGEKKSPPNNANYDKLSR	339
Db	314	SGQQLQMLQLLELLSDSSNSCITWBEETNGEFKMTDDEVARRRNGEKKSPPNNANYDKLSR	373
QY	340	ALARYYDKNINMTKVGRKRYAKEDFHHIAALQPHPESSSMYKYPSDLPMSSVYHAPQK	399
Db	374	ALARYYDKNINMTKVGRKRYAKEDFHHIAALQPHPESSSMYKYPSELPYMSVYHAPQK	433
QY	400	MNFAPARHPALPYTSSSFAPARNPYMNSPPGGITPNTRLAAMHPSILGTYT	451
Db	434	MNFAPARHPALPYTSSSFAPARNPYMNSPGSITPNTRLAPASHSSILGTYT	485

RESULT 7
Q920K9
ID Q920K9 PRELIMINARY; PRT; 462 AA

DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)
DE Erg protein.
GN ERG.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Osawa R., Noguchi H., Taylor T.D., Takeda T., Hattori M., Sakaki Y.;
RT "Mus musculus Erg mRNA.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB073078; BAB69948.1; -.
DR MGD; MGI:95415; Erg.
DR InterPro; IPR000418; Ets.
DR InterPro; IPR002341; HSF-ETS.
DR InterPro; IPR003118; SAM_PWT.
DR Pfam; PF00178; Ets; 1.
DR Pfam; PF02198; SAM_PWT; 1.
DR PROSITE; PS00345; ETS_DOMAIN_1; UNKNOWN_1.
DR PROSITE; PS00346; ETS_DOMAIN_2; UNKNOWN_1.
DR PROSITE; PS0061; ETS_DOMAIN_3; 1.
SQ SQUENCE 462 AA; 52037 MW; 0EA419B53ECA5591 CRC64;

Query Match	88.0%;	Score 2151;	DB 11;	Length 462;
Best Local Similarity	87.6%;	Pred. No. 4.1e-162;		
Matches 402; Conservative	16;	Mismatches 27;	Indels 14;	Gaps 4;

[illegible]

```
Db 189 SYNDLLSLHLHYRETPPLHL---TS--DDVDKALQNSPLMARNTDLYEPPRRSAAW 243
QY 234 TSHSPT-QSKATQPSSTVKTEDQRPOLDPYQILGPTSSRLANPGSGQIQLMQFLEL 292
Db 244 TGHSHLPQSKAAQPSPAVKTEDQRPOLDPYQILGPTSSRLANPGSGQIQLMQFLEL 303
QY 293 LSDSSNSNCITWEGTNGEFTKTDPEVARRWGERKSKPMNMYDKLSRLARYYDKNIMTK 352
Db 304 LSDSSNSNCITWEGTNGEFTKTDPEVARRWGERKSKPMNMYDKLSRLARYYDKNIMTK 363
QY 353 VHGKRYAKPFDFHGLAQLQPHPESSSMYKXPSDLPYMSYHAPPOKNFYAPHPALPY 412
Db 364 VHGKRYAKPFDFHGLAQLQPHPESSSMYKXPSDLPYMSYHAPPOKNFYAPHPALPY 423
QY 413 TSSSFFAAPNPYWSPTGCIYPTNRLPAAHMPSHLGTYY 451
Db 424 TSSSFFAAPNPYWSPTGCIYPTNRLPAAHMPSHLGTYY 462

RESULT 8
Q9PU61 PRELIMINARY; PRT; 451 AA.
AC Q9PU61;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Fli-1 protein.
GN Fli1 OR Fli-1.
OS Brachydanio rerio (zebrafish) (zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20108585; PubMed=10640707;
RA Brown L.A., Rodaway A.R., Schilling T.F., Jowett T., Ingham P.W.,
RA Patient R.K., Sharrocks A.D.;
RT "Insights into early vasculogenesis revealed by expression of the ETS-
RT domain transcription factor Fli-1 in wild-type and mutant zebrafish
RT embryos."
RL Mech. Dev. 90:237-252(2000).
DR EMBL: AJ249590; CAB56832.1; -.
DR HSSP: Q01543; IFLI.
DR ZFIN: ZDB-GENE-980526-426; fli1.
DR InterPro: IPR000418; HSF.
DR InterPro: IPR002341; HSF.ETS.
DR InterPro: IPR003118; SAM_PNT.
DR Pfam: PF00178; Ets; 1.
DR Pfam: PF02198; SAM_PNT; 1.
DR PRINTS: PR00454; ETSDOMAIN.
DR SMART: SM00413; ETS; 1.
DR SMART: SM00251; SAM_PNT; 1.
DR PROSITE: PS00345; ETS_DOMAIN_1; 1.
DR PROSITE: PS00346; ETS_DOMAIN_2; 1.
DR PROSITE: PS50061; ETS_DOMAIN_3; 1.
DR PROSITE: PS50061; ETS_DOMAIN_3; 1.
SQ SEQUENCE 451 AA; 50655 MW; F437D5D08B3549A CRC64;

Query Match 67.0%; Score 1637; DB 13; Length 451;
Best Local Similarity 66.6%; Pred. No. 2e-121;
Matches 303; Conservative 56; Mismatches 88; Indels 8; Gaps 5;
```

```
QY 180 TFSYNADLLSHLYRERGAFTFEPNTSVYPEAQTTRRDLPEQARRSAMTSHSP 239
Db 178 TSVNTEVLLSHLNTLRSSSSISY-NTPSHADQSPRLAAKODASYDAVRRIGMNNHS 236
QY 240 TQSKATQPSSTVKTEDQRPOLDPYQILGPTSSRLANPGSGQIQLMQFLELSDSSNS 299
Db 237 GKSPTVVSQSVSKNPDQRPQDPQYQILGPTSSRLANPGSGQIQLMQFLELSDSSANA 296
QY 300 NCITWEGTNGEFTKTDPEVARRWGERKSKPMNMYDKLSRLARYYDKNIMTKVGRRYA 359
Db 297 GCTWEGTNGEFTKTDPEVARRWGERKSKPMNMYDKLSRLARYYDKNIMTKVGRRYA 356
QY 360 YKFDHGLAQLQPHPESSSMYKXPSDLPYMSYHAPPOKNFYAPHPALPYTSSSFEA 419
Db 357 YKFDHGLAQLQPHPESSSMYKXPSDLPYMSYHAPPOKNFYAPHPALPYTSSSFEA 416
QY 420 APNPYWSPTGCIYPTNRLPA--AHMPSHLGTYY 451
Db 417 PTPYWSPTGCIYPTNRLPA--AHMPSHLGTYY 451

RESULT 9
Q93425 PRELIMINARY; PRT; 432 AA.
AC Q93425;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Fli transcription factor.
GN Fli.
OS Coturnix coturnix japonica (Japanese quail).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Coturnix.
OX NCBI_TaxID=93934;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98358003; PubMed=9694627;
RA Mager A.M., Grabin-Botton A., Ladjali K., Meyer D., Wolff C.M.,
RA Stiegler P., Bonnin M.A., Remy P.;
RT "The avian fli gene is specifically expressed during embryogenesis in
RT a subset of neural crest cells giving rise to mesenchyme."
RL Int. J. Dev. Biol. 42:561-572(1998).
DR EMBL: Y14773; CAAT5077.1; -.
DR EMBL: Y14774; CAAT5078.1; -.
DR HSSP: Q01543; IFLI.
DR InterPro: IPR000418; Ets.
DR InterPro: IPR002341; HSF.ETS.
DR InterPro: IPR003118; SAM_PNT.
DR Pfam: PF00178; Ets; 1.
DR Pfam: PF02198; SAM_PNT; 1.
DR PRINTS: PR00454; ETSDOMAIN.
DR SMART: SM00413; ETS; 1.
DR SMART: SM00251; SAM_PNT; 1.
DR PROSITE: PS00345; ETS_DOMAIN_1; 1.
DR PROSITE: PS00346; ETS_DOMAIN_2; 1.
DR PROSITE: PS50061; ETS_DOMAIN_3; 1.
DR PROSITE: PS50061; ETS_DOMAIN_3; 1.
SQ SEQUENCE 432 AA; 48761 MW; B32596A68CB104EB CRC64;

Query Match 64.6%; Score 1579.5; DB 13; Length 432;
Best Local Similarity 66.7%; Pred. No. 6.7e-117;
Matches 304; Conservative 46; Mismatches 77; Indels 29; Gaps 8;
```

```

Db      118  TTNNRRVIVPADPFLMTQGEHVRQKLENAIKREYGLMEIDTTFEPQMOCKELCKNNKDDFLR 1777
Qy      179  LTPSTNADILLSHLHYLRKRGATITFPNTSYTPAATORTITRPDLPEQARSAMTSHSH 2388
Db      178  TTSLTNTEVLLSHLSYLNRSSSLAY-NTPSHTAASSRLATKEGPPV----- 2233
Qy      239  PTOGKATPQSSSYVKTEDKROPOLPOIIGPTSSRLANGSGOIQLOMFLLELSDSN 2298
Db      224  ----AGTONVKT---TEQAPQDPDPIQIOLPTSSRLANGSGOIQLOMFLLELSDSN 2766
Qy      299  SNCITWECTNGEFPKMTDPEVFAARMGGERKSKPNNNYDKLSRALYYDKDKIMTKVHGKRY 3588
Db      277  ASCITWECTNGEFPKMTDPEVFAARMGGERKSKPNNNYDKLSRALYYDKDKIMTKVHGKRY 3566
Qy      359  AYKDFEHIQALQDPHPRESSMYTPSDLPMSSTYNAHPQKAMFVAHPALPYTSSSF 4188
Db      337  AYKDFEHIQALQDPHPRESSMYTPSDLSMPBYNAHQKRVNVPHPSPMPTSSSF 3666
Qy      419  AAPNPYNSPTGGLYPN---TRLCAPAHMPSHLGYY 451
Db      397  GAASPTWTSFAGSITYPNNPNVFRHNAHVSFHLGSIY 432

```

ID	Q9PUL6	PRELIMINARY:	PRT:	414 AA.
AC	Q9PUL6;			
DT	01-MAY-2000 (TrEMBLrel. 13, Created)			
DT	01-MAY-2000 (TrEMBLrel. 13, last sequence update)			
DT	01-DEC-2001 (TrEMBLrel. 19, last annotation update)			
DE	Filli protein (Fragment).			
GN	Filli.			
OS	Brachydanio rerio (Zebrafish) (Zebra danio).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;			
OC	Cyprinidae; Danio.			
OX	NCBI_TaxId=7955;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=98294174; PubMed=9630750;			
RA	Thompson M.A., Ransom D.G., Pratt S.J., MacLennan H., Kieran M.W.,			
RA	Detrich H.W.I.I., Vail B., Huber P.L., Paw B., Brownlie A.J.,			
RA	Oates A.C., Fritz A., Gates M.A., Amores A., Bahary N., Talbot W.S.,			
RA	Her H., Beler D.R., Postlethwait J.H., Zon L.I.;			
RT	"the cloche and spadetail genes differentially affect hematopoiesis			
RT	and vasculogenesis."			
RT	Dev. Biol. 197;248-269(1998).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RA	Thompson M.A., Ransom D.G., Pratt S.J., MacLennan H., Kieran M.W.,			
RA	Detrich H.W., Vail B., Huber T.L., Paw B., Brownlie A.J., Oates A.C.,			
RA	Fritz A., Gates M.A., Amores A., Bahary N., Talbot W.S., Her H.,			
RA	Beler D.R., Postlethwait J.H., Zon L.I.;			
RL	Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AF17538; AAF01966.1; -			
DR	HSSP; Q01543; 1FLL.			
DR	ZFIN; ZDB-GENE-980526-426; filli.			
DR	InterPro; IPR000448; Ets.			
DR	InterPro; IPR002341; HSF_ETTS.			
DR	InterPro; IPR003118; SAM_PNT.			
DR	Pfam; PF00178; Ets; 1.			
DR	Pfam; PF02198; SAM_PNT; 1.			
DR	PRINTS; PR00454; ETSDOMAIN.			
DR	SMART; SM00413; ETS; 1.			
DR	SMART; SM00251; SAM_PNT; 1.			
DR	PROSITE; PS00345; ETS_DOMAIN_1; 1.			
DR	PROSITE; PS00346; ETS_DOMAIN_2; 1.			
DR	PROSITE; PS50061; ETS_DOMAIN_3; 1.			
FT	NON_TER	1	1	
FT	TER	414	414	
SQ	SEQUENCE	414 AA;	46705 MW;	FE88032B5F6C5C5C1 CRC64;

Query Match	Similarity	60.8%	Score 1486.5	DB 13	Length 414
Best Local Similarity	65.6%	Pred No. 1.5e-109			
Matches 273	Conservative 53	Mismatches 83	Indels 5	Gaps 4	
OY	14	EDOSLFECAY-GSPHLAKTEMTASSSSEYGTGTSKMSPRVPOODWLSQPARVITYKMECPN	72		
DB	1	EDOSLFEPYPAALAPLPIKTDMTASGTQDYGTHNKINIPPOEWIINO--PVRVNVRKREY--	57		
OY	73	NOVNSRNSPDDCCSAVAGSGKMYSSSDNGYCMGNTGSWEKHITPPMNTTERRVIVYADPT	132		
DB	58	DHINSRRESVDCSSGKCKMKMGVTEASOMNTGTWDECCAPPNNNTTERRVIVPADDS	117		
OY	133	LMSTHVRKQMLEMAVKEYGLPPVDLLIFQONIDKEIKCKTKDFOPLTFESYNADILSLH	192		
DB	118	LMSPDHVRQMLDMAIKKEYGLQEIOTDPAFMHSTGDKELCKSKSDPLRLTSVYNTVEVLLSLH	177		
OY	193	HYLRKGAFTFIRPNISVYPELATOITTRBDLPEDQARSAMTSHSHPTQSKATQPOSSSTV	252		
DB	178	NLRLESSSSISV-NTPSHADOSPRLAADDASDYDAVRRFGWNNMNSGSGSPVTVSOVS	236		
OY	253	PYTEQORQOLDPYQILGTPSSRLANPGSQIOLMOLFELLESDSSNSNCITWEGNGEER	312		
DB	237	KNPDQPRRQPDYQILGTPSSRLANPGSQIOLMOLFELLESDSNAGCITWEGNGEER	296		
OY	313	MTDPDEVARRMGEKRSKPMNANTDKLSRALRYYYDKNIMTKVHGKRYAYKEDFHGIAQALQ	372		
DB	297	MTDPDEVARRMGEKRSKPSMANTDKLSRALRYYYDKTMTKVHGKRYAYKIDFHGIAQALQ	356		
OY	373	PHPPSSYAKYVPSDLPYSSSYAHGOKNMFVAHPHALPVSSSFPAANPVPWNSD	428		
DB	357	PHPTSTYAKYVSELPIYPSTYAHGOKNMFVSPHPSPMPTVSSNNEGPTTPYWNSD	412		

Query Match	39.5%	Score 965;	DB 4;	Length 476;
Best Local Similarity	45.7%	Pred. No. 3,4e-68;		
Matches 218;	Conservative 43;	Mismatches 130;	Indels 86;	Gaps 13;
7	EALSVSHDGLFECAVGSPLHAKTEMTASS-----SEYCO-----TSKMSPRV	51		
1	:	:	:	:
2	:	:	:	:
3	:	:	:	:
4	:	:	:	:
5	:	:	:	:
6	:	:	:	:
7	:	:	:	:
8	:	:	:	:
9	:	:	:	:
10	:	:	:	:
11	:	:	:	:
12	:	:	:	:
13	:	:	:	:
14	:	:	:	:
15	:	:	:	:
16	:	:	:	:
17	:	:	:	:
18	:	:	:	:
19	:	:	:	:
20	:	:	:	:
21	:	:	:	:
22	:	:	:	:
23	:	:	:	:
24	:	:	:	:
25	:	:	:	:
26	:	:	:	:
27	:	:	:	:
28	:	:	:	:
29	:	:	:	:
30	:	:	:	:
31	:	:	:	:
32	:	:	:	:
33	:	:	:	:
34	:	:	:	:
35	:	:	:	:
36	:	:	:	:
37	:	:	:	:
38	:	:	:	:
39	:	:	:	:
40	:	:	:	:
41	:	:	:	:
42	:	:	:	:
43	:	:	:	:
44	:	:	:	:
45	:	:	:	:
46	:	:	:	:
47	:	:	:	:
48	:	:	:	:
49	:	:	:	:
50	:	:	:	:
51	:	:	:	:
52	:	:	:	:
53	:	:	:	:
54	:	:	:	:
55	:	:	:	:
56	:	:	:	:
57	:	:	:	:
58	:	:	:	:
59	:	:	:	:
60	:	:	:	:
61	:	:	:	:
62	:	:	:	:
63	:	:	:	:
64	:	:	:	:
65	:	:	:	:
66	:	:	:	:
67	:	:	:	:
68	:	:	:	:
69	:	:	:	:
70	:	:	:	:
71	:	:	:	:
72	:	:	:	:
73	:	:	:	:
74	:	:	:	:
75	:	:	:	:
76	:	:	:	:
77	:	:	:	:
78	:	:	:	:
79	:	:	:	:
80	:	:	:	:
81	:	:	:	:
82	:	:	:	:
83	:	:	:	:
84	:	:	:	:
85	:	:	:	:
86	:	:	:	:
87	:	:	:	:
88	:	:	:	:
89	:	:	:	:
90	:	:	:	:
91	:	:	:	:
92	:	:	:	:
93	:	:	:	:
94	:	:	:	:
95	:	:	:	:
96	:	:	:	:
97	:	:	:	:
98	:	:	:	:
99	:	:	:	:
100	:	:	:	:
101	:	:	:	:
102	:	:	:	:
103	:	:	:	:
104	:	:	:	:
105	:	:	:	:
106	:	:	:	:
107	:	:	:	:
108	:	:	:	:
109	:	:	:	:
110	:	:	:	


```

RP SEQUENCE FROM N.A.
RA Boquet-Muchemled B., Leroux R., Fontaine F.;
RT "Isolation of members of ets gene family in polychaete annelids.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ276692; CAB99437.1; -.
DR HSSP; Q01543; 1F1I.
DR InterPro; IPR000418; Ets.
DR InterPro; IPR002341; HSF_ETs.
DR Pfam; PF00178; Ets; 1.
DR PRINTS; PR00454; ETSDOMAIN.
DR SMART; SM00413; ETS; 1.
DR PROSITE; PS50061; ETS_DOMAIN_3; 1.
FT NON_TER 1
FT NON_TER 160
SQ SEQUENCE 160 AA; 18035 MW; A088F1EB570897AB CRC64;

Query Match 22.2%; Score 543.5; DB 5; Length 160;
Best Local Similarity 65.6%; Pred. No. 1.8e-35;
Matches 107; Conservative 17; Mismatches 30; Indels 9; Gaps 6;

QY 287 QFLELLSDSSNSNCITWEGTNGEFKMTDPEVARRWGERKSKP--NMNYDKLSRALRYX 344
DB 1 QFLELLSNSANSHITWEGTNGEFKLVDPDEFARRMGARKSKPNMNMNYDKLSRALRYX 60
QY 345 YDKNMTKYHGKRYAKKPFDFHGTAQALOPHPRESSMYKYPSPDLPYMSYHAHPQKNFYA 404
DB 61 YDKNMTKYHGKRYAKKPFDFHGTAQALOPHPRESSMYKYPSPDLPAAYKYOODI-LMSGYH-HTSKLNLMA 118
QY 405 PHPALPVTSSEFFAAPNPYWNPSPTG-GIYPNTRLPAAHMPSH 446
DB 119 AAHP-MASSASGFFPPSPYSSPAGSNLXPNT---SNHMSH 157

RESULT 15
ID Q22355 PRELIMINARY; PRT: 377 AA.
AC Q22355;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-JAN-1999 (TREMBLrel. 09, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Hypothetical 41.5 kDa protein.
GN T08H4.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE-90069613; PubMed-9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Steillys L.;
RT "The sequence of C. elegans cosmid T08H4.";
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; U39470; AAC71120.1; -.
DR HSSP; Q01543; 1F1I.
DR InterPro; IPR000418; Ets.
DR InterPro; IPR002341; HSF_ETs.
DR Pfam; PF00178; Ets; 1.
DR PRINTS; PR00454; ETSDOMAIN.
DR SMART; SM00413; ETS; 1.

```

```

DR PROSITE; PS00345; ETS_DOMAIN_1; 1.
DR PROSITE; PS00346; ETS_DOMAIN_2; 1.
DR PROSITE; PS50061; ETS_DOMAIN_3; 1.
KW Hypothetical protein.
SQ SEQUENCE 377 AA; 41466 MW; D78A6BFA2472D409 CRC64;

Query Match 21.4%; Score 524; DB 5; Length 377;
Best Local Similarity 51.9%; Pred. No. 2.1e-33;
Matches 120; Conservative 14; Mismatches 30; Indels 67; Gaps 10;

QY 226 EQARRSAMTSHS-----PQSK-----ATQPSSTVKTED 257
DB 110 DQSRQFTYESSNSGNGAATSGSSSTESKSDVNISMNAAPAGS---KSED 165
QY 258 QR-POL-----DPYQILGPTSSRLANPGSGQIQLMQFLELLS 294
DB 166 HNPSPFMNLSSYYTGALKLSNSTSFANPPYQILGPTSKNLHSGSGQIQLMQFLELLS 225
QY 295 DSSNSNCITWEGTNGEFKMTDPEVARRWGERKSKPNMNYDKLSRALRYXDKNMTKYH 354
DB 226 DKRYSEVITWEGTNGEFKLVDPDEFVARRWGERKSKPNMNYDKMSRALRYXDKNIMAKYH 285
QY 355 GKRYAYKPFDFHGTAQALOPHPRESSMYKYPSPDLPYMSYHAHPQKNFYAP 405
DB 286 GKRYAYKPFDFHGTAQALQ--PPTAS--HPQD--YFNS-HA---MGRAP 324

Search completed: November 9, 2002, 16:53:25
Job time : 58.8288 secs

```